

UNIVERSITE DU QUEBEC

MEMOIRE

PRESENTE A

L'UNIVERSITE DU QUEBEC A TROIS-RIVIERES

COMME EXIGENCE PARTIELLE

A L'OBTENTION

DE LA MAITRISE EN PSYCHOLOGIE

PAR

FRANCE MIMEAULT

PREDICTION DU RENDEMENT ACADEMIQUE A L'UNIVERSITE

A PARTIR D'UN TEST PSYCHOMETRIQUE

MESURANT LA CAPACITE D'APPRENTISSAGE

AVRIL 1985

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

Table des matières

Introduction	1
Chapitre premier - Contexte théorique et expérimental	5
Origine de l'emploi des tests en contexte de sélection académique	6
Validité prédictive et différentielle de certains tests psychométriques	11
Sources de variation des coefficients de validité	33
Objectifs et hypothèses	41
Chapitre II - Description de l'expérience	50
Sujets	51
Examen psychométrique	52
Mesure du rendement scolaire à l'université	56
Déroulement de l'expérience	58
Chapitre III - Analyse des résultats	60
Méthodes d'analyse	61
Présentation des résultats	62
Interprétation des résultats	86
Résumé et conclusion	106
Appendice A - Examen psychotechnique	113
Appendice B - Résultats complémentaires	124

Remerciements	132
Références	133

Sommaire

Le but de la présente recherche consiste à évaluer la validité prédictive d'un instrument récemment élaboré (Henning, 1976), le Learning Ability Profile (LAP), qui permet d'obtenir une mesure quantitative de la capacité d'apprentissage. En second lieu et à titre exploratoire, cette étude s'intéresse à la validité différentielle de ce test ainsi qu'à l'influence de certains facteurs susceptibles de faire varier les coefficients de validité.

La procédure consiste à évaluer la relation entre les scores au LAP et les résultats académiques ultérieurs d'un groupe expérimental composé de 45 étudiants inscrits en première année au Baccalauréat en psychologie à l'Université du Québec à Trois-Rivières. L'ampleur de la covariation entre les résultats au test et les notes académiques fut quantifiée par l'analyse corrélationnelle du produit des moments de Pearson.

Les coefficients de corrélation obtenus révèlent qu'il n'y a pas de relations significatives au seuil de

0,05 entre les résultats au LAP et la moyenne cumulative de deux sessions universitaires. Par contre, lorsque les scores à ce test ont été mis en relation avec les résultats à des cours individuels qui font partie intégrante de la moyenne cumulative, on observe des corrélations positives et significatives pour trois cours sur une possibilité de dix.

Les cours universitaires les mieux prédits par les résultats au LAP semblent avoir deux caractéristiques communes. D'une part, on remarque que ce sont les cours correspondant à des matières scolaires à caractère plus scientifique que verbal qui obtiennent les plus forts coefficients de validité prédictive, ce qui confirme la valeur potentielle de ce test en tant que prédicteur différentiel du rendement académique. D'autre part, on observe que les coefficients de validité prédictive significatifs ont été obtenus dans les cours pour lesquels on relève une grande variabilité dans les notes attribuées par les enseignants. Cette dernière constatation nous incite à considérer le faible pouvoir discriminant de certaines évaluations du corps professoral comme l'une des principales variables pouvant expliquer la faiblesse relative de certains coefficients de corrélation obtenus.

Bien que les présents résultats n'apportent qu'un appui mitigé concernant la valeur de prédiction du LAP, des études de validation supplémentaires s'imposent avant d'en arriver à des conclusions définitives sur la pertinence de son utilisation en contexte de sélection ou d'orientation scolaire.

Introduction

Le problème du pronostic du rendement scolaire occupe, en psychologie appliquée, une place fort importante. Aux Etats-Unis par exemple, plusieurs revues scientifiques publient dans chacun de leur numéro un ou des articles touchant cette question. Malgré le nombre considérable de recherches réalisées ici et ailleurs, malgré la diversité des instruments utilisés et des types de mesure considérés jusqu'à présent à des fins de classement et de sélection, le bilan dressé par Garrett en 1949 concernant la valeur relative de ces instruments en termes de validité de pronostic semble encore valable aujourd'hui: les résultats scolaires antérieurs demeurent le meilleur moyen de prédiction du rendement académique ultérieur; viennent ensuite les résultats à des tests d'habiletés mentales ou d'aptitudes spécifiques. Sauf exception, c'est ce que les auteurs observent au niveau universitaire où le problème du pronostic se pose avec une acuité particulière principalement dans les programmes d'études contingentés.

Afin d'expliquer l'échec partiel des examens d'admission pour la sélection des étudiants universitaires, certains chercheurs (Chansky, 1964; Etaugh et al., 1972;

Lavin, 1965; Michael et al., 1983; Motoyama et Wolins, 1980) ont remis en cause la fidélité et la validité des notes scolaires lorsque celles-ci sont utilisées à titre de critères alors que d'autres auteurs ont directement accusé la valeur des tests psychométriques, remettant par le fait même en cause la pertinence de leur utilisation (Baby, 1975; Boucher, 1982; Roy, 1972; Thacker et Williams, 1974). En conséquence, les recherches en ce domaine ont pris deux orientations distinctes: certains chercheurs se sont acharnés à valider d'autres critères du rendement scolaire alors que d'autres recherches ont été consacrées à la validation de nouveaux examens psychotechniques.

La présente étude s'inscrit dans le deuxième groupe de recherches puisqu'elle cherche à vérifier l'existence d'une relation significative entre les résultats à un test d'apprentissage, le Learning Ability Profile, et les résultats scolaires chez un groupe d'étudiants de première année en psychologie à l'Université du Québec à Trois-Rivières.

Au plan du déroulement, cette recherche respecte les trois étapes majeures suivantes. Dans un premier temps, afin de préciser la problématique ayant initié cette étude, le contexte théorique et expérimental expose la

pertinence d'utiliser les tests d'aptitudes en contexte de sélection en milieu éducationnel. La deuxième partie de ce chapitre présente quelques recherches américaines et canadiennes qui se sont intéressées à la validité prédictive et différentielle de certains tests psychométriques. La troisième partie s'intéresse aux facteurs qui font varier l'ampleur des coefficients de validité prédictive obtenus à partir du même prédicteur. Enfin, les objectifs et les hypothèses de cette recherche sont présentés dans la dernière partie de ce chapitre.

Dans le second chapitre, l'échantillon choisi sera décrit, les instruments de mesures utilisés à titre de prédicteur et de critère seront présentés, ainsi que le protocole expérimental suivi. Finalement, les méthodes d'analyse ainsi que la présentation et l'interprétation des résultats font l'objet du dernier chapitre, lequel sera résumé dans la conclusion.

Chapitre premier

Contexte théorique et expérimental

Dans ce chapitre, nous aborderons la problématique de la prédiction du rendement académique faite à partir de tests psychométriques. Pour ce faire, nous exposerons les fondements théoriques qui sont à l'origine de l'emploi des examens d'admission dans les collèges et les universités. Ce même chapitre présentera quelques recherches empiriques qui se sont intéressées à la validation de prédicteurs du succès scolaire. Nous terminerons ce chapitre par l'énumération des facteurs qui font varier la grandeur des coefficients de validité prédictive ainsi que par la présentation des objectifs et des hypothèses que cette recherche se propose de vérifier.

Origine de l'emploi des tests en contexte de sélection académique

La nécessité des tests en contexte de sélection académique découle principalement de la connaissance de trois éléments de réalité (Boucher, 1975). Le premier de ces éléments est l'existence du phénomène du contingentement qui peut se définir comme étant une limitation du nombre d'étudiants qu'une institution décide d'admettre dans un programme donné. Selon Cormier (1978), les rai-

sons communément invoquées pour contingenter un programme d'étude gravitent autour des mêmes thèmes: le manque de ressources humaines et matérielles, le manque de lieux de stage et la situation du marché du travail. Qu'on soit d'accord ou non avec cette politique, il semble qu'on ne peut guère y échapper, les universités ayant reconnu qu'elles ne pourront agir différemment avant plusieurs années puisque la hausse des inscriptions dans certains programmes pourrait se poursuivre dans les années 80 (Statistique Canada, 1982). Cette hausse des inscriptions s'explique par un accroissement constant du nombre de femmes parmi les étudiants du post-secondaire, une hausse des exigences en matière d'études pour occuper un emploi et un taux de chômage moins élevé chez les diplômés universitaires comparativement à d'autres groupes de jeunes moins scolarisés. De plus, la pratique du contingentement se justifie par le fait que la vraie démocratisation de l'enseignement supérieur ne consiste pas à rendre cet enseignement disponible à tout le monde mais consiste plutôt à le rendre disponible à tous ceux qui possèdent les qualités intellectuelles, la préparation et la motivation pour oeuvrer efficacement dans ce domaine exigeant qu'est celui de l'université (Gaudry, 1982).

Le second élément à l'origine des tests découle du choix de l'excellence comme principal critère d'admission. Au lieu de se définir des critères d'admission reliés à l'âge, au sexe, au collège, au niveau socio-économique, à la région ou à l'occupation des parents, les institutions universitaires ont préféré définir des critères d'admission qui prennent en considération l'excellence des candidats, la valeur et la capacité de ceux-ci. Cette mesure se justifie par le fait que dans une situation de sélection en contexte universitaire, l'admission d'un étudiant qui échouera en cours de route entraîne une perte de temps et d'argent considérable pour l'institution et pour l'étudiant, tandis que le rejet d'un candidat qui aurait réussi prive la société d'un membre utile et l'étudiant d'une position méritoire (Ewen, 1969). D'autre part, bien que la législation québécoise sur les droits et libertés de la personne ne concerne pas directement la sélection en contexte académique, elle concrétise une orientation irréversible vers une plus grande objectivité dans tous les processus de sélection.

Le troisième élément à l'origine des tests découle des insuffisances du dossier scolaire (Anastasi, 1982; Cronbach, 1970; Linn, 1982). L'absence d'examens

uniformes au niveau collégial de même que des différences au niveau de la présentation des résultats rendent difficile l'utilisation véritable de l'information contenue dans le dossier académique d'un candidat. Pour pallier à cette difficulté, on cherche habituellement à mettre au point des systèmes de conversion susceptibles de rendre comparables les résultats des divers étudiants. En dépit de la somme considérable d'efforts qu'elles nécessitent souvent, les diverses initiatives en ce sens demeurent nécessaires certes, mais très imparfaites et leurs lacunes profondes risquent le plus souvent de remettre en question les principes fondamentaux d'équité, de justice sociale et d'efficacité sous-jacents à tout travail de sélection (Boucher, 1972). En effet, cette tâche de sélection ne consiste pas à choisir les meilleurs dossiers scolaires et à mettre de côté les plus faibles, mais plutôt à départager un certain nombre d'étudiants à partir d'un grand nombre de candidatures présentant des dossiers équivalents.

Les tests psychométriques ont donc précisé pour objectif premier de compléter les renseignements du dossier scolaire par des informations objectives, communes, précises et standardisées sur les ressources des candidats afin d'être en mesure de faire une évaluation plus juste

des chances de succès de ces derniers. Cependant, le fait d'utiliser de tels tests ne peut garantir l'atteinte de cet objectif. Pour qu'un examen d'admission soit valable, voir même utile, il se doit d'être valide. La validité prédictive d'un test, c'est-à-dire la mesure du degré de relation entre des prédicteurs (caractéristiques nécessaires pour effectuer adéquatement la tâche) et des critères (indices de rendement d'une personne dans la situation envisagée) est la qualité primordiale à considérer pour l'introduction d'un test psychométrique dans un processus de sélection (Voyer, 1972).

Il existe plusieurs types d'instruments qui sont susceptibles de fournir ces informations, on retrouve notamment: les tests d'acquisitions scolaires, les tests d'intérêts et de motivation, les inventaires de personnalité et les tests d'aptitudes et de fonctionnement intellectuel. Toutefois, dans le cadre de la présente recherche, nous ne retiendrons pour fin d'analyse que les tests d'aptitudes et de fonctionnement intellectuel. Le problème du pronostic du rendement scolaire effectué à partir de tests ayant donné lieu à un nombre très considérable de travaux et de recherches sur des prédicteurs et des critères différents, seuls seront considérés ici les travaux nécessaires à situer dans leur contexte théorique et expérimental les

objectifs que la présente recherche se propose d'atteindre.

Validité prédictive et différentielle
de certains tests psychométriques

Un des plus anciens tests psychométriques élaborés à des fins d'admission et d'orientation scolaire est le College Board Scholastic Aptitude Test (SAT). Le SAT fut développé en 1926 par le College Entrance Examination Board qui est un organisme à but non lucratif composé de 2 500 représentants du monde de l'éducation. Ce test est administré dans plus de 48% des collèges américains en vue d'évaluer le potentiel des étudiants du niveau high school à poursuivre des études universitaires (Hargadon, 1981).

La première forme du SAT comprenait neuf sous-tests. En 1929, les auteurs de la batterie décident qu'il était devenu nécessaire de diviser celle-ci en deux sections distinctes: l'une mesurant l'aptitude verbale (SAT-V) et l'autre l'aptitude pour les mathématiques (SAT-M). Le SAT-V comprend 90 items qui se composent d'antonymes, de synonymes, de phrases à compléter et de questions sur la compréhension d'un texte. Pour sa part, le SAT-M comprend 60 items composés de problèmes d'arithmétique et d'algèbre

ainsi que de concepts géométriques élémentaires. Pour chaque question, l'évalué doit choisir entre cinq possibilités de réponses et il a trois heures pour compléter ce test.

Il est à noter que la division du SAT en deux sections distinctes avait pour but de pronostiquer de façon différentielle les probabilités de réussite des candidats en fonction de la nature du collège et du programme d'étude choisis (Angoff, 1971). En plus des scores verbal et numérique, les formes récentes de ce test fournissent des mesures d'achievement dans différents domaines d'études. Contrairement aux mesures d'aptitudes qui sont utilisées pour prédire le rendement éventuel dans une activité quelconque, les tests d'achievement cherchent à mesurer le niveau actuel de compétence ou les connaissances dans diverses matières scolaires (Adams, 1964).

Un nombre considérable de recherches ont été effectuées afin de valider cet examen psychotechnique, les résultats tendant à démontrer qu'il existe une relation substantielle entre le score au SAT et les résultats académiques qui sont généralement exprimés sous forme de moyennes cumulatives pour la première année universitaire

(Astin, 1971; Houston, 1983; Linn, 1982; Goldman et Slaughter, 1976). Une synthèse de plus de 700 études sur la validité du SAT effectuée par Ford et Campos (1977: voir Slack et Porter, 1980) révèle que les résultats actuels sont similaires à ceux obtenus en 1926: la validité prédictive du SAT pour des groupes composés de sujets des deux sexes est en moyenne de 0,40 pour le SAT-V, de 0,35 pour le SAT-M et de 0,42 au niveau de la corrélation multiple, ce qui améliore de 0,08 la prédiction du rendement scolaire faite à partir des seuls résultats au high school. Ces coefficients de validité moyens ne doivent toutefois pas être considérés comme étant des valeurs absolues puisque des corrélations négatives sont parfois observées entre les scores au SAT et la moyenne cumulative (Linn et Dunbar, 1982).

Les résultats de cette synthèse sont interprétés différemment selon les auteurs; pour certains (Hartnett et Felmesser, 1980; Nairn, 1980), le SAT ne prédit pas mieux le rendement scolaire qu'une paire de dés alors que pour d'autres (Dubois, 1972; Goldman et Slaughter, 1976; Linn, 1982), les résultats obtenus avec ce test constituent une augmentation appréciable de la prédiction du succès scolaire.

D'autres études se sont intéressées plus particulièrement à la validité différentielle du SAT. Le calcul des coefficients de corrélation entre les deux sous-tests représente une première étape dans l'étude de la validité différentielle. La moyenne des coefficients de corrélation obtenus entre le score verbal et le score numérique est supérieur à 0,60 ce qui est beaucoup trop élevé pour des tests qui prétendent mesurer des aptitudes différentes (Fricke, 1965; Wallace, 1972). La deuxième étape pour évaluer la validité différentielle consiste à établir des corrélations entre les résultats à chacun des sous-tests et le rendement dans différentes matières scolaires. On devrait s'attendre à ce que les résultats au SAT-V soient plus fortement corrélés avec les résultats aux cours à caractère verbal tels que l'anglais et l'histoire alors qu'on devrait obtenir une corrélation plus forte entre les résultats au SAT-M et ceux obtenus dans les cours à caractère scientifique. Or très peu de recherches permettent d'étudier ce genre de relations et les résultats sont contradictoires. Une étude effectuée en 1976 par Goldman et Slaughter révèle que le SAT-V prédit le rendement dans les cours de psychologie et de sociologie avec plus d'efficacité qu'il ne le fait pour les cours de biologie, de chimie et de physique; d'autres recherches (Passons, 1967;

Dubois, 1972) démontrent que le score verbal est supérieur au score numérique dans la prédiction du succès dans dix cours différents dont l'histoire, la psychologie, les mathématiques et les sciences physiques.

Les Achievement Tests du SAT sont quelque peu différents des tests d'aptitudes (SAT-V et SAT-M) parce qu'ils sont conçus dans le but de mesurer les apprentissages réalisés dans des matières scolaires spécifiques. Etant donné leur nature, on est en droit de s'attendre à une corrélation élevée entre les résultats à un test particulier et le rendement au cours correspondant et par ailleurs, à une corrélation faible entre le même test et le rendement aux autres cours. Toutefois, les faits tendent plutôt à réfuter cette hypothèse. Fricke (1965) rapporte les résultats de trois études montrant que, de façon générale, les tests destinés à mesurer des apprentissages spécifiques n'ont pas une valeur de pronostic supérieure aux autres tests mesurant des aptitudes.

Suite à ces observations, cet auteur conclut que le SAT-V, le SAT-M et les Achievement Tests ne sont pas utiles pour prédire de manière différentielle le rendement au niveau universitaire. Chacun des tests semble

évaluer l'aptitude scolaire en général plutôt qu'une aptitude spécifique.

Devant le succès mitigé obtenu avec le SAT, de nombreux chercheurs se sont acharnés à développer et à valider d'autres tests psychométriques. Ces recherches ont eu comme résultat la création de diverses batteries de tests d'aptitudes et d'achievement pour des fins de sélection en milieu éducationnel universitaire. Le ACT Assessment Program (ACT) fut créé à cette fin en 1959 par le American College Testing Program et est administré dans plus de 2 600 collèges et universités américaines (Evarts, 1980). Le ACT est constitué d'un test d'intérêt et de quatre tests d'aptitudes scolaires mesurant la capacité prévisible d'effectuer un travail dans quatre domaines d'études distincts soit: l'anglais, les mathématiques, les sciences sociales et les sciences naturelles.

Le English Usage Test comprend 75 items qui évaluent la compréhension de l'étudiant pour les conventions de l'anglais écrit ainsi que l'usage des éléments de base tels que la ponctuation, la grammaire, le style et l'organisation des phrases. Ce test a une durée de 40 minutes.

Le Mathematics Usage Test est constitué de 40 items et a une durée de 40 minutes. Il mesure le raisonnement mathématique de l'étudiant et contient des problèmes d'arithmétique, d'algèbre, de géométrie et de statistique.

Le Social Studies Reading Test et le Natural Sciences Test contiennent chacun 52 items et l'évalué a 35 minutes pour compléter chacun des sous-tests. Ils évaluent la compréhension, le raisonnement analytique et les habiletés à la résolution de problèmes requis pour chacun de ces deux domaines scolaires.

La validité du ACT est comparable à celle du SAT. En général, ces deux tests sont interchangeables pour prédire la moyenne cumulative des étudiants au niveau de la première année universitaire (Linn, 1982; Hills, 1978; Wallace, 1972). Une étude synthèse d'Astin (1971) auprès de 180 institutions scolaires rapporte des coefficients de validité moyens de 0,35 pour les sujets masculins et de 0,43 pour les sujets féminins entre la moyenne cumulative de deux sessions universitaires et les résultats au ACT. Il faut cependant noter que les résultats obtenus avec ce test psychométrique varient d'une recherche à l'autre. Ainsi, un autre examen des données accumulées

sur la validité prédictive du ACT (ACT Standard Research Service, 1964: voir Munday, 1970) relève des indices de validité qui s'échelonnent d'un coefficient négatif près de zéro à un coefficient positif très élevé.

Il semble par ailleurs que les tests d'aptitudes scolaires du ACT ne possèdent pas les attributs nécessaires pour en faire un prédicteur différentiel. Premièrement, de forts coefficients de corrélation sont observés entre les différents sous-tests ce qui suggère que tous les sous-tests de cette batterie feraient appel à des aptitudes intellectuelles similaires (Engelhart, 1965; Hills, 1978). Deuxièmement, étant donné le fait que les différents sous-tests du ACT sont conçus pour mesurer des aptitudes dans des matières scolaires spécifiques, on est en droit de s'attendre à une corrélation élevée entre les résultats à un test particulier et le rendement au cours correspondant et par ailleurs à un faible coefficient de validité entre le même test et le rendement aux autres cours. Les faits tendent cependant à infirmer cette hypothèse; une étude de Findley (1965) portant sur 705 étudiants universitaires rapporte un coefficient de corrélation de 0,496 entre le sous-test Social Studies Reading Test et les résultats au cours de sciences naturelles alors qu'il observe

un coefficient de 0,457 entre les résultats à ce même cours et le sous-test Natural Sciences Reading Test. Finalement, il apparaît que l'on pourrait n'employer que le sous-test pour lequel on obtient les plus forts coefficients de validité pour prédire le succès académique puisque l'addition des autres sous-tests du ACT n'augmentent l'efficacité de la prédiction que de façon négligeable (Wallace, 1972).

La pratique de l'évaluation des aptitudes des candidats pour les fins d'admission au niveau universitaire s'est étendue à l'admission aux études de deuxième et de troisième cycles dès 1937 avec l'apparition du Graduate Record Examination National Program (GRE). Cette batterie de test comprend un test d'aptitudes (GRE-AT) en plus de 20 Advanced tests dont le contenu est en relation avec des matières académiques acquises antérieurement.

Le test d'aptitudes du GRE a une conception similaire à celui du SAT mais les questions ont un niveau de difficulté beaucoup plus élevé que ce dernier. Comme celui-ci, l'évalué a trois heures pour compléter l'épreuve qui est constituée d'une section verbale comportant des synonymes, des antonymes et des analogies, ainsi que d'une section quantitative où l'étudiant est confronté à des

problèmes d'algèbre, de statistique et d'interprétation de graphiques.

Sa validité prédictive est difficilement quantifiable de façon absolue puisque celle-ci varie considérablement d'une recherche à l'autre. Par exemple, un examen de la documentation effectué en 1968 par le Educationnal Testing Service (Lanholm et al., 1968: voir Littlepage et al., 1978) révèle que les coefficients de corrélation entre d'une part le GRE-V, le GRE-Q et le Advanced Test in Psychology et d'autre part, les résultats académiques aux études supérieures en psychologie varient d'un coefficient de corrélation positif modéré (0,35, 0,45, 0,35 respectivement) à un coefficient de corrélation négatif (-0,28, -0,24, -0,35 respectivement). Certains travaux (Bean, 1975; Hackman et al., 1970) rapportent même que les résultats scolaires antérieurs ne corrèlent pas toujours de façon significative avec les résultats académiques aux études graduées.

Afin d'expliquer les faibles coefficients de validité prédictive obtenus dans certaines recherches s'intéressant au GRE, certains auteurs ont remis en cause l'utilisation de la moyenne cumulative comme indice du

rendement scolaire aux études supérieures. Dans le but de résoudre ce problème, on a essayé de valider d'autres critères tels que l'obtention du doctorat (Willingham, 1974), le temps requis pour l'obtention du doctorat (Hirschberg et Itkin, 1978) ou par une évaluation de la qualité de la thèse produite par l'étudiant (Porter et Wolfe, 1975). Les résultats de ces recherches ne sont malheureusement guère plus encourageants (Hartnett et Willingham, 1980; Thacker et Williams, 1974).

Dans un autre ordre d'idée, Carter (1972) s'est intéressé plus particulièrement à la validité différentielle du GRE en analysant les données de 18 recherches de validation de cette batterie. Il observe que, de façon générale, les étudiants inscrits dans les secteurs à caractère hautement scientifique tels que les mathématiques, la chimie et la physique, réussissent mieux que les étudiants en lettres le sous-test quantitatif. Cependant, il remarque aussi que les mêmes étudiants du secteur scientifique surpassent les étudiants en lettres au niveau du rendement au facteur verbal. Carter arrive à la conclusion que les décisions relatives à l'admission seraient biaisées si elles étaient faites à partir d'un score global du GRE ou à partir d'une seule mesure verbale ou quantitative.

Bien qu'au Canada il n'existe pas de programmes nationaux d'évaluation et de sélection des étudiants universitaires, des efforts en ce sens ont déjà été faits notamment avec la création du Service d'admission au Collège et à l'Université (SACU). Le SACU est un organisme interprovincial à but non lucratif regroupant les collèges, les universités, les ministères de l'éducation et d'autres membres intéressés. Selon Elliott (1972), cet organisme a été créé afin de préparer des tests acceptables et adaptables pour éviter une balkanisation au Canada qui obligerait l'étudiant à passer un test à chaque université où il présente une demande d'admission et pour éviter une politique très coûteuse de sélection: celle qui se fait après la première année universitaire.

Les tests du SACU comprennent des tests rédigés dans les deux langues officielles. Ils comportent entre autre le Test d'aptitude aux études post-secondaire (TAGEPS) ou Canadian Scholastic Aptitude Test (CSAT) et le Test de français langue maternelle (TFLM) ou Canadian English Language Achievement Test (CELAT). Ces tests ont une conception à peu près similaire à ceux du SAT mais ils ne contiennent que des items canadiens (Lamontagne, 1972). Le TAGEPS ou CSAT comporte deux scores: un pour la section

verbale et l'autre pour la section mathématique, tandis que le TFLM ou CELAT n'en comporte qu'un seul. Les trois scores ainsi obtenus, exprimés sous forme numérique, indiquent la position qu'occupe un étudiant sur une échelle de 200 à 800 avec une moyenne de 500 et un écart-type de 100. Ces tests n'ont pas de notes de passage, les universités utilisant les scores à leur gré. La valence relative des scores du SACU par rapport à la moyenne scolaire dans l'équation de prévision de réussite va de 10 à 35%; la valence médiane est d'environ 25%, soit une partie pour les scores de ces tests contre trois parties pour la moyenne scolaire (Elliott, 1972).

La validité prédictive des tests du SACU est difficile à établir puisqu'il semble n'y avoir que très peu de recherches qui s'y sont intéressées. On peut toutefois mentionner l'étude de Harvey effectué en 1972 au Nouveau-Brunswick dont les résultats tendent à démontrer que le CSAT et le CELAT sont des prédicteurs aussi fidèles et discriminateurs que le SAT l'est pour les élèves américains. Cette étude relève des coefficients de corrélation de l'ordre de 0,431, 0,468 et 0,469 entre les différents tests du SACU (CSAT-V, CSAT-M et CELAT respectivement) et la moyenne de différents cours de première année

universitaire pour un groupe de sujets anglophones. Pour le groupe de sujets francophones, les coefficients de validité prédictive sont: 0,457 pour le TAGEPS-V, 0,224 pour le TAGEPS-M et 0,281 pour le TFLM.

Une autre recherche effectuée cette fois au Cégep de la Pocatière (Roy, 1972) rapporte des corrélations de 0,318 pour le TAGEPS-V, de 0,039 pour le TAGEPS-M et 0,388 pour le TFLM. Les résultats de cette recherche sont interprétés par l'auteur comme n'apportant rien de valable dans le processus de sélection alors que pour Lavoie (1973), les résultats obtenus sont enviabiles si on les compare à d'autres études s'intéressant à la prédiction du rendement académique.

En milieu québécois, l'utilisation de tests en vue de prédire le rendement académique ne constitue pas un phénomène nouveau puisque dans les années 40, la Fédération des Collèges Classiques administrait à un nombre important de candidats au cours classique de tels examens psychométriques. Dans cette optique, les travaux de Gadbois et Brault (1960) montrent qu'il existe une relation significative entre le quotient intellectuel évalué à partir de l'Examen Lauzon et les notes scolaires. D'une part, ils

remarquent que les résultats scolaires médiocres et les départs du collège se situent parmi les élèves dont le quotient intellectuel est inférieur à la médiane du groupe et d'autre part, que les corrélations calculées entre les quotients intellectuels et les notes académiques varient de 0,24 ($p > 0,05$) à 0,34 ($p < 0,005$) selon les années d'études.

Gadbois et Brault (1960) voulaient également évaluer la validité différentielle de l'Epreuve individuelle d'intelligence générale (Barbeau-Pinard) chez un groupe d'étudiants inscrits en Philo II. A partir de ce test, chaque élève se voyait attribuer trois quotients intellectuels différents: un quotient intellectuel verbal, basé sur ses réponses à plusieurs séries de questions (dialogue avec l'examineur); un quotient intellectuel non-verbal, basé sur ses réponses à plusieurs séries de problèmes où les éléments visuels et moteurs sont au premier plan; enfin, un quotient intellectuel global, basé sur l'ensemble du test. L'hypothèse sous-jacente à cette partie de l'étude était qu'il est logique de s'attendre à ce que ces divers quotients intellectuels aient des rapports différents avec le succès dans les différentes disciplines, ce qui implique de la validité différentielle. Par exemple

le succès en philosophie, qui suppose une bonne capacité d'abstraction, serait en relation avec un fort quotient verbal tandis que le quotient intellectuel non-verbal, basé sur un jugement plus concret et sur diverses manipulations, serait lié au succès en laboratoire.

Les résultats empiriques de cette étude ne confirment que partiellement l'hypothèse voulant que le Barbeau-Pinard soit un prédicteur différentiel. Ainsi, le succès au cours de dissertation philosophique semble lié à un quotient intellectuel global, et le succès en physique, à un haut quotient intellectuel non-verbal. Quant au succès dans l'ensemble des matières de Philo II, il est en relation avec les divers types de quotients intellectuels: une corrélation de 0,39 ($p < 0,005$) est observée avec le quotient intellectuel global alors que des coefficients de corrélation de l'ordre de 0,33 ($p < 0,01$) et 0,27 ($p < 0,05$) sont obtenus avec le quotient non-verbal et verbal respectivement.

Toujours dans cette perspective d'améliorer le pronostic du succès académique en milieu universitaire, le Registraire de l'université de Montréal de concert avec le Ministère de l'éducation du Québec a créé les Tests d'Aptitudes aux Etudes Universitaires. Cette batterie de tests, communément appelé TAEU, a connu un essor considéra-

ble dans les années 1970 de sorte qu'elle est actuellement administrée à tous les étudiants francophones du Québec qui font une demande d'admission dans certaines facultés de l'Université de Montréal et de l'Université de Sherbrooke (Goulet, 1982). Les TAEU ont pour objectif principal de compléter les renseignements du dossier scolaire par des informations communes, précises, objectives et standardisées sur les ressources des candidats afin de permettre une évaluation plus juste des chances de succès de ces derniers (Boucher, 1975).

Cette batterie comprend 12 tests originaux, standardisés et normalisés auprès des étudiants universitaires francophones. Ces tests visent à évaluer cinq aptitudes relativement générales soit: la compréhension verbale, le raisonnement, l'ingéniosité, l'habileté perceptuelle et la mémoire. L'étude de la corrélation entre la performance à chacun des tests et la performance à chacun des cours de première année dispensés dans tous les départements est effectuée à chaque année scolaire. Ces études se poursuivent jusqu'à l'établissement d'une équation de régression par matière et pour tous les cours d'un même département.

Pour les fins de sélection, un classement est effectué de tous les individus qui ont fait une demande d'admission dans un département donné. Ce classement s'opère de la façon suivante: un score est calculé pour chaque candidat à partir de l'équation de régression qui sert à prédire la réussite dans l'ensemble des cours du département en question, puis les individus sont ordonnés par rapport à leur score. Les responsables de l'admission retiennent d'abord tous les candidats admissibles en fonction des autres critères et rejettent tous ceux qui ne satisfont pas aux exigences d'admission. Si un choix s'avère nécessaire, parmi les candidats présentant des dossiers équivalents, la préférence pourra être donnée à ceux qui obtiennent le meilleur score aux tests d'aptitudes.

Les résultats d'études tendent à démontrer qu'il existe une relation positive entre les résultats aux TAEU et la réussite à l'université. Ceci implique que les étudiants qui, dans un programme donné, présentent les meilleurs résultats aux tests sont en général ceux qui réussissent le mieux leurs études et inversement ceux qui subissent des échecs dans un programme d'études ont généralement obtenu les résultats les plus faibles aux tests. Ainsi Belzil (1980) rapporte un coefficient de corrélation

de 0,38 ($p < 0,01$) entre le rendement aux TAEU et le rendement à cinq cours du Certificat général en droit. Guédon (1979) relève pour sa part un coefficient de validité de 0,288 ($p < 0,05$) entre les résultats académiques du Cégep et la moyenne générale de dix cours de première année en médecine dentaire; en combinant les résultats aux tests d'aptitudes avec les résultats scolaires au Cégep, le coefficient de validité prédictive passe de 0,288 à 0,375 ($p < 0,01$).

Dans le cadre d'un mémoire déposé en 1977, Cardinal s'est intéressée à la validité différentielle des TAEU et dégage quatre constatations majeures de sa recherche. Une première constatation est que les divers programmes d'études font surtout intervenir une dimension relativement générale que l'on peut nommer "fonctionnement intellectuel global". Les divers aspects mesurés tels que la compréhension verbale, le raisonnement, l'habileté perceptuelle, l'ingéniosité et la mémoire interviennent dans tous les programmes d'études. Une seconde constatation générale de cette recherche est qu'il semble faux d'affirmer que tel programme d'études fait appel à tel mode de fonctionnement: la compréhension verbale intervient autant dans les études de la Faculté des Sciences que dans celles de la Faculté de Lettres. Une troisième constata-

tion révèle la présence de scores totaux caractéristiques pour des groupes de programmes d'études: en considérant l'ensemble des résultats aux TAEU d'un candidat, on peut prédire avec un pourcentage de succès de l'ordre de 84% son appartenance ou non à un secteur scientifique. La dernière constatation a trait au phénomène de compensation d'une aptitude par une autre: un résultat faible à un test n'est pas un indice certain d'un faible niveau de rendement scolaire si ce candidat a par ailleurs des résultats élevés à d'autres tests. Cependant, si tous les résultats aux TAEU sont faibles, on est alors beaucoup plus certain que le rendement scolaire ultérieur sera plus faible.

Récemment, deux autres recherches se sont consacrées à la prédiction du succès scolaire en milieu universitaire québécois, mais cette fois en faisant intervenir la notion d'apprentissage comme prédicteur. L'une de ces recherches réalisée par Boucher (1982) voulait quantifier la relation entre une mesure de la vitesse d'apprentissage et la moyenne cumulative de première année universitaire pour un groupe de 78 étudiants au département de psychologie de l'Université de Montréal. L'épreuve expérimentale consistait en un texte de 1 300 mots traitant d'anthropologie à partir duquel les candidats

avaient à répondre à 22 questions de compréhension et de connaissances et ce, en quatre essais successifs de vingt minutes. La tâche des sujets consistait à répondre correctement à ces questions dans un nombre minimum d'essais tout en maintenant l'exactitude des réponses dans les essais subséquents. Les résultats de cette recherche révèlent que le rythme d'apprentissage présente une corrélation de 0,43 ($p < 0,01$) avec les résultats scolaires alors que le résultat cumulatif, qui ne tient pas compte de la vitesse d'apprentissage, a une corrélation de 0,51 ($p < 0,01$) avec ce même critère.

Pour sa part, Fabi (1983) s'est intéressé au pronostic du rendement académique à partir d'une mesure de la capacité d'apprentissage réalisée avec le Learning Ability Profile (Henning, 1976). Ce test comprend, dans sa version modifiée, 68 items à niveau de difficulté croissant qui se composent d'analogies de figures et de symboles, d'analogies arithmétiques et séries numériques, d'analogies et séries de lettres et finalement, d'analogies de mots. Le score à ce test constitue un indice quantitatif qui reflète la capacité de raisonnement inductif et déductif aussi bien que les habiletés cognitives à la résolution de problèmes. Cette étude révèle un coefficient de cor-

rélation de 0,312 ($p < 0,01$) entre les résultats à cet examen psychotechnique et la moyenne cumulative de trois sessions universitaires chez un groupe de 80 étudiants inscrits à un cours au département d'administration et d'économie à l'Université du Québec à Trois-Rivières. Selon Fabi, bien que ce test semble être un prédicteur valide du rendement académique, certains facteurs ont influencé de façon négative l'ampleur des coefficients de corrélation. C'est pourquoi celui-ci suggère qu'il serait pertinent de répéter une telle démarche de validation en impliquant cette fois un échantillon plus homogène où l'on pourrait recueillir des résultats académiques obtenus dans les mêmes cours, à partir des mêmes critères et des mêmes modes d'évaluation.

L'énumération qui précède n'est pas exhaustive mais elle permet de constater la prolifération des études s'intéressant à la prédiction du rendement académique. Dans l'ensemble, les résultats de ces recherches tendent à démontrer l'utilité d'avoir recours à des tests psychométriques pour prédire la moyenne cumulative ou les notes à certains cours universitaires. On remarque également que pour des tests utilisés à très grande échelle comme le SAT, le ACT et le GRE, les résultats obtenus en terme

de coefficients de validité prédictive varient considérablement d'une recherche à l'autre. Considérant la diversité des résultats relevés dans les diverses recherches de validation, certains auteurs (Albright et al., 1963; Ghiselli, 1966; Guion, 1965; Zeleznik et al., 1983; Houston, 1983) recommandent que des procédures de validation empiriques soient effectuées dans chaque institution scolaire ainsi que pour chaque programme d'étude puisque des généralisations au sujet d'un test quant à son pouvoir prédictif sont essentiellement impossibles.

Une telle variation au niveau des coefficients de validité prédictive obtenus pour un même test a provoqué l'élaboration d'une série de recherches suggérant une gamme considérable de facteurs explicatifs. La prochaine section sera donc consacrée à la présentation des principaux facteurs responsables d'une telle variation.

Sources de variation des coefficients de validité

Récemment, un certain nombre de recherches (Burke, 1984; Linn et al., 1981; Pearlman et al., 1980; Schmidt et al., 1982) se sont employées à identifier les sources de variations de l'ampleur des coefficients de validité prédictive observés pour un test psychométrique

utilisé en contexte de sélection. Ces recherches ont permis de déceler sept sources de variation à l'origine d'une telle fluctuation: (1) les erreurs d'échantillonnage causées par la taille de l'échantillon, (2) des différences entre les études au niveau de la fidélité du critère, (3) des différences entre les études au niveau du degré de dispersion des résultats au prédicteur, (4) des différences entre les études au niveau de la fidélité du prédicteur, (5) les erreurs typographiques ou de calcul, (6) des différences au niveau du critère utilisé et (7) des différences dans la structure du test employé.

Parmi les sept sources de variance mentionnées, les auteurs suggèrent de porter une attention particulière aux trois premières puisqu'elles sont responsables de plus de 70% des fluctuations observées au niveau de la grandeur des coefficients de validité prédictive obtenus pour un test particulier.

La première source de variation a trait au petit nombre de sujets impliqués dans certaines recherches de validation en contexte scolaire. Comme le soulignent Callender et Osburn (1980), un échantillon trop restreint permet l'influence de certains facteurs non contrôlés sur

la variable dépendante. Selon la direction que prend cette influence, impondérable, la valeur de la relation entre la variable indépendante et la variable dépendante peut être tantôt sous-estimée tantôt sur-estimée.

La deuxième source de variance importante se rapporte aux différences observées entre les études au niveau de la fidélité du critère utilisé. Ce facteur n'est pas à négliger puisque comme le souligne Cronbach (1970): "When a test fails to predict a rating, it is hard to say whether this is the fault of the test or the rating" (p. 127).

La faiblesse de certains coefficients de fidélité associés aux mesures du rendement académique est imputable à deux sources d'imprécisions. L'une de ces sources réside dans le fait que les barèmes d'évaluation varient d'un professeur à l'autre; certains enseignants ne tiennent compte que du niveau de connaissance acquis par l'étudiant alors que d'autres attribuent aussi des points à la participation orale aux cours, à l'attitude en classe, au degré d'intérêt manifesté par l'étudiant et parfois à des caractéristiques individuelles qui ne sont aucunement en relation avec le rendement scolaire (Chansky, 1964; Etaugh et al., 1972).

L'autre source d'imprécision fait référence au caractère non discriminant de certaines évaluations du corps professoral où l'on remarque une forte propension à attribuer des cotes A et B. Ce phénomène d'inflation des notes scolaires semble avoir pris une envergure considérable. Une enquête effectuée par Suslow (1976: voir Singleton et Smith, 1978) auprès de 15 universités américaines révèle qu'entre 1963 et 1974, la moyenne cumulative des étudiants est passée de 2,49 à 2,94, les cotes A et B constituant 71% des notes attribuées en 1974. Pour certains auteurs (Davidson, 1975; White, 1975) une telle pratique d'évaluation ne correspond aucunement aux différences réelles qui existent dans le niveau d'accomplissement des étudiants, elle s'explique plutôt par le fait que les enseignants ont abdiqué devant leur rôle d'évaluateur en appliquant des standards moins élevés pour apaiser les étudiants. En contexte de validation d'un prédicteur, une faible variabilité des résultats au critère ne peut qu'avoir une influence négative sur l'ampleur des coefficients de validité prédictive (Dole et Baggaley, 1979; Hartnett et Willingham, 1980; Merenda et Reilly, 1971; Michael et al., 1983; Thacker et Williams, 1974).

Le troisième facteur explicatif relié aux va-

riations des coefficients de validité obtenus pour un test particulier réside dans des différences au niveau du degré de dispersion des résultats au prédicteur. Une limitation de l'étendue des scores au test psychométrique est fréquemment observée dans ce genre de recherche et ce phénomène semble attribuable en grande partie au fait qu'en contexte universitaire, une sélection graduelle est déjà intervenue pour éliminer les moins doués ce qui provoque une faible dispersion d'aptitudes au sein du groupe de sujets étudiés. En général, plus un groupe d'individus est homogène par rapport aux traits mesurés, plus la corrélation est faible (Dayhaw, 1979).

L'effet de la restriction des scores au niveau du prédicteur (range restriction effects) sur l'ampleur des coefficients de validité prédictive a été mis particulièrement en évidence par Schrader (1977: voir Linn, 1982) dans un sommaire de plus de 726 études de validité sur le Law School Admission Test (LSAT). Celui-ci a montré qu'il existait une corrélation de 0,58 entre l'ampleur des écarts-types au LSAT et les coefficients de validité prédictive obtenus en contexte académique. Ce lien entre l'écart-type des scores au test et le coefficient de validité ne s'applique pas uniquement à ce test psychométrique puisqu'une autre recherche de cet auteur (Schrader, 1971)

révèle, pour le Scolastic Aptitude Test (SAT), un coefficient de validité médiant de 0,44 quand l'écart-type à ce test est de 85 ou plus alors que lorsque la valeur de l'écart-type est inférieure à 75, le coefficient de validité médiant est de 0,31.

Toujours dans l'optique d'expliquer les variations de l'ampleur des coefficients de validité prédictive relevés pour un test d'admission aux études universitaires, d'autres types de variables ont été proposés. Dans un article intitulé Factors influencing the predictability of college grades, Munday (1970) rapporte que 42% de la variance est attribuable à certaines caractéristiques des institutions scolaires impliquées dans ce genre de recherche et à des différences dans la composition de l'échantillon pour lequel les corrélations sont calculées.

Au niveau des caractéristiques institutionnelles, cette étude révèle que le pronostic du rendement académique se fait avec plus de précision dans les universités et les collèges qui exercent un plus grand contrôle sur la vie de l'étudiant. Ce contrôle aurait une influence positive sur la motivation de l'étudiant et favoriserait l'émergence d'un climat propice aux études. Plus spécifiquement, les

données de cette recherche démontrent que les institutions qui obtiennent les plus forts coefficients de validité prédictives sont: des institutions privées, des universités où le nombre d'étudiants admis en première année à plein temps ne dépasse pas 500 et des institutions où une grande partie des étudiants réside sur le campus.

En ce qui concerne les différences au niveau de la composition de l'échantillon, la recherche de Munday révèle que les coefficients de validité prédictive les plus élevés sont obtenus dans les études impliquant des groupes de sujets pour lesquels on relève: premièrement, des résultats au test psychométrique qui témoignent d'une grande dispersion d'aptitudes (écart-type élevé); deuxièmement, des résultats moyens au test psychométrique supérieurs à la moyenne nationale et troisièmement, dans les études où l'on observe un fort pourcentage de sujets féminins. Bien que les deux premiers facteurs se rapportent aux tests psychométriques, ils sont indépendants puisque les résultats tendent à démontrer que lorsque la moyenne des scores au prédicteur est élevée, l'écart-type des résultats à ce même prédicteur est faible.

La dernière source de variance reliée à la com-

position de l'échantillon se rapporte aux différences remarquées entre l'ampleur des coefficients de validité prédictive obtenus chez des groupes de sujets féminins comparativement aux groupes de sujets masculins. De nombreuses recherches supportent l'existence d'une telle différence qui s'établit en faveur des sujets féminins (Astin, 1971; Jones, 1970; Khan, 1973; Lavin, 1965; Stanley, 1967; Seashore, 1962). L'étude synthèse d'Astin (1971) auprès de 180 institutions scolaires révèle par exemple des coefficients de corrélation de 0,35 pour les sujets masculins ($N=19\ 524$) et de 0,43 pour les sujets féminins ($N=17\ 057$) entre la moyenne cumulative de deux sessions universitaires et les résultats au ACT. En plus des différences significatives ($p < 0,01$) au niveau des coefficients de validité, cette recherche relève aussi des différences significatives qui vont dans le même sens au niveau de la moyenne cumulative et des résultats moyens au test psychométrique. Selon Astin, pour expliquer cette inégalité entre les sexes, il faut songer aux motivations propres à chaque groupe et non à des différences d'ordre intellectuel. Il semble qu'aux études universitaires, les filles sont plus axées sur le succès scolaire proprement dit alors que les garçons canalisent leurs énergies dans plusieurs activités compétitives telle que le sport.

Khan (1973) estime pour sa part que de telles différences sont reliées aux attentes de la société et au processus de sélection des étudiants dans les universités. Il suggère qu'une partie des garçons qui vont à l'université le font par conformisme ou tout simplement pour ne pas être en contradiction avec les attentes de leurs parents, ce qui entraîne une faible réalisation de leurs potentiels. Par contre, les filles auraient des objectifs bien définis en entrant à l'université et prendraient les moyens adéquats pour les atteindre. Comme autre explication, Khan avance la possibilité que les comités de sélection dans les universités soient plus favorable à l'admission des postulants masculins de sorte que les sujets féminins acceptés seraient de loin de meilleurs candidats.

Ce bref exposé sur les sources de variation des coefficients de validité obtenus avec des tests d'admission aux études universitaires nous amène maintenant à introduire les objectifs et les hypothèses que la présente recherche se propose de vérifier.

Objectifs et hypothèses de la recherche

Henning (1976) soutient que le Learning Ability

Profile (LAP) permet d'obtenir une appréciation du raisonnement inductif et déductif chez les sujets ainsi que de leurs habiletés à résoudre des problèmes. Par conséquent, cet auteur suggère que ce test psychométrique peut s'avérer un outil utile sur le plan de l'orientation et de la sélection dans un milieu de formation. La seule recherche empirique qui, actuellement, s'est intéressée à la validité prédictive de ce test est celle de Fabi (1983) qui rapporte un lien corrélational significatif ($p < 0,01$) entre le rendement au LAP et la moyenne cumulative d'étudiants inscrits en deuxième année au baccalauréat en administration. Ce résultat amène Fabi à inférer la validité du LAP en ce qui a trait à la prédiction du rendement scolaire en milieu universitaire et à recommander la répétition d'une telle démarche de validation afin d'établir des normes plus précises quant aux qualités métrologiques de ce test.

La présente étude s'inscrit dans cette perspective de recherche puisqu'il s'agit d'évaluer la validité prédictive du LAP en relation avec les résultats scolaires chez un groupe de 45 étudiants de première année inscrits au département de psychologie à l'Université du Québec à Trois-Rivières.

En se référant à la classification de Lavin (1965) qui est schématisée au tableau 1, la présente recherche se situe principalement dans la première catégorie (A) de ce modèle puisque l'objectif fondamental consiste à évaluer la validité de pronostic d'un instrument mesurant la capacité d'apprentissage, le Learning Ability Profile, en mettant en relation un indice global de la capacité d'apprentissage et la moyenne cumulative de deux sessions universitaires. Il s'agit d'un processus de validation en fonction d'un critère bien défini (Anastasi, 1982; Bélanger, 1977; Campbell, 1960) où l'on vérifie dans quelle mesure ce critère, le rendement académique au niveau universitaire, peut être adéquatement prédit à partir des résultats à cet examen psychotechnique.

La première hypothèse de cette recherche s'énonce donc ainsi: "Il existera une relation significative entre les scores au LAP et la moyenne cumulative de deux sessions universitaires.

Comme objectif secondaire et à titre purement exploratoire, cette étude vise à évaluer la validité différentielle du LAP. Il s'agit ici d'évaluer dans quelle mesure ce test peut prédire avec plus d'acuité les résul-

Tableau 1

Typologie des recherches mettant en relation les aptitudes
et le rendement académique (Lavin, 1965)

Prédicteur	Critère	
	Global	Multidimensionnel
Global	(A) Mesure globale d'une aptitude utilisée pour prédire la moyenne cumulative	(B) Mesure globale d'une aptitude utilisée pour prédire les résultats à des cours spécifiques
Multidimensionnel	(C) Plusieurs mesures de dimensions spécifiques d'une aptitude utilisées pour prédire la moyenne cumulative	(D) Plusieurs mesures de dimensions spécifiques d'une aptitude utilisées pour prédire les résultats dans des cours spécifiques

tats à certains types de cours (catégorie B du modèle de Lavin). Il doit être noté qu'aucune évidence empirique n'est disponible concernant la possibilité que ce test soit un prédicteur différentiel du succès scolaire. Rappelons aussi que les résultats obtenus à partir de tests comme le SAT, le ACT et le GRE sont équivoques quant à la possibilité de prédire d'une façon différentielle la réussite dans des cours spécifiques.

L'hypothèse qui va suivre se justifie donc essentiellement de la façon suivante: ce prédicteur étant composé en grande partie d'items qui font appel à des aptitudes dites non-verbales ou quantitatives, on peut s'attendre à ce que les scores au LAP soient plus fortement corrélés avec les résultats aux cours à caractère scientifique comparativement aux cours à prédominance verbale. Bien qu'une telle dichotomie soit fréquemment rencontrée dans la documentation (Carter, 1972; Goldman et Slaughter, 1976; Passons, 1967), aucune des recherches consultées ne propose une définition de ces deux termes. Dans le cadre de cette recherche, un cours à caractère scientifique sera défini comme étant un cours qui se caractérise par la présentation et l'évaluation de données objectives et de connaissances exactes alors qu'un cours à prédominance verbale correspondra

à une matière scolaire où les connaissances enseignées sont sujettes à discussion, à interprétation et où il est souvent possible de donner plusieurs réponses à un même problème. L'hypothèse en rapport avec cet objectif s'énonce ainsi: "Les scores au LAP seront plus fortement corrélés avec les résultats à des cours universitaires à caractère scientifique qu'avec ceux obtenus pour les cours à prédominance verbale".

En dernier lieu, cette démarche expérimentale veut évaluer l'effet de certaines variables sur l'ampleur des coefficients de validité prédictive qui seront obtenus à partir du LAP. La première variable étudiée se rapporte à l'influence de la grandeur de l'écart-type au niveau du test psychométrique sur la grandeur du coefficient de validité prédictive. Elle s'inspire d'une documentation récente (Burke, 1984; Linn et al., 1981; Pearlman et al., 1980; Schmidt et al., 1982) qui démontre qu'il existe une relation positive entre cette mesure du degré de variation des scores au prédicteur et les coefficients de validité obtenus.

L'étude de cette variable est motivée par le fait qu'un écart-type relativement faible au niveau des scores au LAP a été observé dans l'étude de Fabi (1983),

cet auteur ayant soupçonné ce facteur d'avoir eu un effet atténuateur sur l'ampleur des coefficients de validité observés. Bien qu'une telle restriction des résultats au prédicteur est habituellement expliquée par une sélection graduelle qui s'est effectuée à l'admission au collège et à l'université (Linn et Dunbar, 1982), dans le cas du LAP un autre facteur pourrait être à l'origine de cette faible dispersion des scores.

Ce facteur résiderait dans la méthode de correction utilisée (answer-until-correct) qui consiste pour le sujet à continuer à répondre à chacune des questions jusqu'à l'obtention de la bonne réponse. Par la suite un score de 0, 1, 3, 4, 6, 7, ou 9 sur 10 est attribuée et cela, même si le sujet n'a pas répondu correctement à la question au premier essai. Il nous semble que la possibilité d'attribuer autant de points pour des réponses initialement fausses peut avoir un effet négatif sur la grandeur de l'écart-type des résultats au test. Pour pallier à ce problème, la présente recherche se propose de faire appel à une autre méthode de correction en plus de celle qui vient d'être brièvement décrite. Cette deuxième méthode de correction, beaucoup plus traditionnelle, est nommée inferred number right score (Hanna, 1977) et consiste à ne considérer que le premier choix de réponses des

sujets. Si la réponse initiale de l'évalué correspond à la bonne réponse, un score de 10 points est attribué pour cette question alors que si l'évalué ne découvre pas la solution à son premier essai, aucun point ne lui est alloué pour cet item. Le détail concernant ces deux méthodes de correction sera présenté au deuxième chapitre dans la section décrivant l'instrument psychométrique utilisé.

En prenant pour acquis que cette deuxième méthode de correction permettra éventuellement d'obtenir un écart-type plus élevé au niveau des scores au LAP, l'hypothèse suivante est formulée: "Les coefficients de validité prédictive obtenus avec le LAP à partir de la méthode de correction inferred number right score (INR) seront significativement plus élevés que ceux relevés à partir de la méthode answer-until-correct (AUC)".

La deuxième variable étudiée en rapport avec les facteurs susceptible d'avoir un effet sur l'ampleur des coefficients de validité prédictive concerne la fidélité du critère. Plus spécifiquement, elle a trait au caractère non discriminant de certaines évaluations du corps professoral où l'on observe une forte propension à attribuer des cotes A et B comme résultat scolaire. Selon certains auteurs (Dole et Baggaley, 1979; Hartnett et

Willingham, 1980; Merenda et Reilly, 1971; Michael et al., 1983; Thacker et Williams, 1974), une telle pratique d'évaluation qui se caractérise par une faible variabilité des résultats au critère ne peut qu'avoir une influence négative sur l'ampleur des coefficients de validité prédictive. Considérant l'importance de ce facteur, l'hypothèse suivante est formulée: "Il existera une relation significative entre l'ampleur de l'écart-type des résultats à un cours universitaire et la grandeur du coefficient de validité prédictive obtenu pour ce cours".

Enfin, la dernière variable retenue pour fin d'analyse concerne les différences relevées dans la documentation scientifique entre l'ampleur des coefficients de validité prédictive obtenus chez les sujets féminins comparativement à ceux obtenus par les sujets masculins (Astin, 1971; Jones, 1970; Khan, 1973; Lavin, 1965). Les différences remarquées dans ces recherches s'établissant en faveur des sujets féminins, il est donc permis de formuler cette hypothèse: "Les coefficients de validité prédictive obtenus chez les sujets féminins seront plus élevés que ceux relevés pour les sujets masculins et ce, pour l'ensemble des cours considérés".

Chapitre II

Description de l'expérience

Ce second chapitre est consacré à la méthodologie employée pour valider le LAP en contexte académique. Dans la première partie, l'échantillon des sujets impliqués dans cette recherche est décrit. Dans la seconde, les instruments de mesure au prédicteur et au critère sont présentés en précisant les qualités métrologiques de ces variables. Ensuite, la partie protocolaire expose la manière selon laquelle l'expérience a été réalisée.

Sujets

Le groupe expérimental est composé de 45 étudiants inscrits en première année au programme de baccalauréat en psychologie à l'Université du Québec à Trois-Rivières. Le groupe est constitué de 24 femmes et de 21 hommes dont l'âge moyen est de 22 ans. Tous ont participé à cette recherche sur une base volontaire malgré le fait qu'ils ont été sollicités durant une période normale de cours.

Soulignons que le choix d'impliquer des étudiants provenant d'un seul domaine d'étude s'explique par le sou-

ci d'avoir un échantillon homogène à partir duquel on peut recueillir des résultats scolaires provenant de cours communs où l'évaluation des connaissances acquises se fait à partir des mêmes barèmes et des mêmes modes d'évaluation.

Examen psychotechnique

Le score obtenu pour chaque étudiant au Learning Ability Profile (LAP) constitue un indice quantitatif qui reflète la capacité de raisonnement inductif et déductif aussi bien que les habiletés cognitives à la résolution des problèmes. Autrement dit, le succès à ce test dépend de la capacité à restructurer des relations et à découvrir les principes qui leur confèrent une signification. Ces habiletés sont considérées par l'auteur (Hanning, 1976) comme des déterminants importants en ce qui concerne les possibilités d'apprentissage de nouvelles tâches et cela, aussi bien en contexte académique qu'organisationnel.

La version utilisée de ce test psychométrique comprend 68 questions à niveau de difficulté croissant qui se distribuent comme suit: 31 analogies de figures et de symboles, 20 analogies arithmétiques et séries numériques, 15 analogies et séries de lettres et 2 analogies de mots (originellement 14). Il est à noter que la réduc-

tion du nombre d'items découle de l'élimination de 12 questions dont la résolution exige la connaissance de la langue anglaise. Toutefois, cette réduction respecte le rationnel de l'auteur de cette échelle qui avait prévu l'utilisation de ce test chez des groupes ethniques non-anglophones. Dans cette perspective, elle a élaboré un mode de pondération qui permet d'obtenir un score total sur 800 même pour les sujets n'ayant répondu qu'à 68 questions. Ce mode de pondération consiste à attribuer à chacune des questions non répondues les mêmes scores que ceux obtenus pour des items présentant des degrés de difficulté équivalents. Dans le cadre de la présente recherche, deux scores au test seront donc considérés pour chacun des sujets: un résultat sur 680 qui constitue le score brut obtenu par l'évalué et un score sur 800 qui tient compte du mode de pondération proposé par Henning (1976).

Ce test se caractérise par le fait qu'il est en lui-même une situation d'apprentissage puisque la feuille réponse utilisée est du type answer-until-correct carbon erasure procedure (AUC), cette méthode permet de donner un feedback immédiat concernant l'exactitude de la réponse (Hanna, 1977). Pour chacune des questions, le sujet doit

choisir entre quatre possibilités de réponses. Une fois son choix effectué, il doit gratter la tache d'encre sous la réponse choisie et il verra apparaître un des quatre scores suivants: 10 pour la bonne réponse, -6 pour la plus médiocre, -3 et -1 pour les deux autres. Les directives indiquent que l'on doit nécessairement découvrir la bonne réponse pour chacune des questions avant de passer à l'item suivant même si cela exige de gratter les quatre taches d'encre pour y parvenir. L'étudiant peut ainsi déterminer immédiatement son résultat pour chaque question en effectuant la somme algébrique des scores découverts, le maximum possible étant de 10 et le minimum de 0. L'appendice A rapporte le détail de la consigne et présente les items du test.

Une telle procédure permet donc au sujet d'apprendre certains principes de raisonnement à mesure qu'il effectue le test (Hanna, 1977). Cette caractéristique explique probablement le très fort pourcentage de satisfaction rapporté par Henning (1976) lors de ses expériences de validation, ce qui a été confirmé lors d'entrevues informelles réalisées à la fin de la présente expérimentation. Selon Fabi (1983), la motivation intrinsèque provoquée par la passation de cet instrument de mesure constitue un

avantage non négligeable pour des utilisations éventuelles à des fins de sélection ou d'orientation où l'on observe trop souvent un niveau d'intérêt mitigé chez les sujets subissant certains tests papier-crayon traditionnels. De plus, cet auteur suggère que le fait que ce test ne soit pas chronométré pourrait contribuer à diminuer le stress, souvent responsable d'un faible rendement en situation d'évaluation chez certains sujets.

Dans la présente étude, un autre mode de correction a été utilisé en plus de celui proposé par l'auteur du LAP. Cette deuxième méthode consiste à ne considérer que le premier choix de réponses des sujets pour chacun des 68 items. Si la réponse initiale de l'évalué correspond à la bonne réponse, un score de 10 points est attribué pour cette question alors que si l'évalué ne découvre pas la solution à son premier essai, aucun point ne lui est alloué pour cet item. Cette méthode de correction, beaucoup plus traditionnelle (Hanna, 1977), est nommée inferred number right score (INR). Elle a été introduite afin de vérifier l'influence possible d'une faible dispersion des résultats au prédicteur sur l'ampleur des coefficients de validité prédictive (hypothèse 2).

Le manuel de normes du LAP révèle des coefficients de fidélité interne de 0,95 (items pairs vs impairs) et de 0,93 (40 premiers vs 40 derniers) calculés à partir de la formule split-half de Spearman-Brown. Au niveau de la validité, aucun coefficient n'est présenté, le seul argument présenté par l'auteur (Henning, 1966) consistant à souligner la relation positive observée entre les moyennes de résultats au test et le niveau de scolarité. Cependant, une recherche effectuée à l'Université du Québec à Trois-Rivières chez un groupe d'étudiants en administration révèle une relation significative ($p < 0,01$ et $r = 0,312$) entre les résultats au LAP et la moyenne cumulative d'une année universitaire (Fabi, 1983).

Mesure du rendement scolaire à l'université

En guise de mesure du rendement académique, nous avons retenu la moyenne cumulative à l'hiver 1983 ainsi que les résultats pour dix cours suivis lors de cette même année académique. Il convient de souligner que la moyenne cumulative et les résultats à des cours individuels sont fréquemment employés à titre de critère dans la documentation scientifique consacrée à la validation de tests psychométriques dans différents milieux éducationnels. Bien

que la fidélité de ces critères varie d'une recherche à l'autre, elle est qualifiée d'acceptable par de nombreux chercheurs (Chansky, 1964; Goldman et Slaughter, 1976; Linn et al., 1981; Schmidt et Kaplan, 1971).

Rappelons que la moyenne cumulative est calculée à la fin de chaque session à partir de toutes les notes obtenues dans le programme d'étude, en attribuant une valeur numérique aux lettres de la notation littérale et en pondérant par le nombre de crédits accordés pour chacun des cours. Pour en arriver à cette dernière, on attribue aux lettres les valeurs numériques suivantes: A=4, B=3, C=2, D=1 et E=0. Cette moyenne cumulative peut varier entre 0 et 4 et est calculée à la deuxième décimale.

Les cours individuels de première année universitaire retenus dans cette recherche sont: Psychologie de la perception I (PEX-1002), Psychologie de l'apprentissage (PEX-1005), Histoire de la psychologie (PPS-1002), Psychophysiologie I (PSP-1001), Statistiques en psychologie (STT-1014), Psychologie dynamique I (PCL-1010), Psychologie des groupes (PCL-1013), Psychophysiologie II (PSP-1002), Psychologie génétique I (PSD-1006) et Psychologie sociale I (PLS-1004).

Déroulement de l'expérience

L'examen psychotechnique Learning Ability Profile (LAP) fut administré au début de la session d'automne 1982 lors d'une seule séance d'expérimentation effectuée à l'intérieur du cours Psychologie de l'apprentissage. Il est à noter que les étudiants étaient libres de participer à cette expérience. Bien que le LAP est un test non chronométré, une période maximale de trois heures était allouée pour compléter l'épreuve. Cette limite de temps a été fixée en tenant compte de l'expérimentation effectuée par Fabi (1983) et elle s'est avérée suffisante pour la totalité des sujets.

La présente étude est présentée aux sujets comme étant une recherche de maîtrise en psychologie dont l'objectif est d'évaluer dans quelle mesure un test d'apprentissage peut prédire le rendement académique futur. En retour de leur participation, la confidentialité des résultats obtenus leur est assurée.

Suite à cette présentation, le matériel se rapportant au test est distribué et la consigne est lue aux étudiants en insistant sur les particularités de la feuille

réponse et des symboles utilisés dans l'exercice (voir appendice A). Afin de familiariser les sujets avec la tâche à accomplir, deux exemples de problèmes prévus à cette fin dans le questionnaire et la feuille réponse sont présentés aux sujets avant que ne débute la séance d'expérimentation proprement dite.

Dans un deuxième temps, les résultats scolaires obtenus par les sujets durant l'année suivant la passation du LAP ont été relevés à partir du bulletin officiel émis par l'établissement d'enseignement. Une fois ces données recueillies, elles sont soumises à une analyse de corrélation afin de quantifier la relation entre le score au test et les notes académiques.

Chapitre III

Analyse des résultats

Avant de présenter les résultats proprement dits et de les interpréter, il s'impose de décrire brièvement les différentes méthodes d'analyse statistique employées dans cette recherche.

Méthodes d'analyse

L'analyse des résultats fait appel à divers indices statistiques qui se retrouvent dans la programmation informatique du Statistical Package for the Social Sciences (SPSS) élaborée par Nie et al. (1975). Etant donné que le schéma expérimental implique essentiellement l'évaluation du niveau de corrélation entre une variable indépendante et des variables dépendantes continues (Kerlinger, 1973), la corrélation dite du moment des produits a été utilisée et cet indice statistique a été obtenu à l'aide du programme Pearson Corr dans la programmation SPSS.

Il convient de souligner que plusieurs recherches récentes de validation psychométrique rapportent une méthode d'analyse similaire et ce, aussi bien en contexte

québécois (Belzil, 1980; Boucher, 1982; Guédon, 1979) qu'en contexte américain (Baldauf, 1978; Camp et Clawson, 1979; Costin, 1978; Omizo et Rivera, 1980).

L'analyse de ces coefficients de corrélation s'effectuera à partir de leur seuil de signification respectif. Bien que la fixation d'un tel seuil constitue une démarche arbitraire et restrictive aux yeux de certains auteurs (Henkel, 1976), particulièrement dans des recherches comme celle-ci où il n'existe pas de procédure rationnelle permettant d'attribuer précisément un coût à l'erreur alpha ou à l'erreur bêta, il n'en demeure pas moins qu'une convention expérimentale fixe ce seuil à 0,05 (Selltiz et al., 1977). Il est en effet convenu de rejeter l'hypothèse nulle quand l'analyse statistique indique que la différence ou la relation observée ne résulterait pas du simple hasard plus de cinq fois sur cent. Cette convention sera respectée dans le cadre de la présente étude.

Présentation des résultats

L'exposé des résultats se divise en deux parties distinctes: une première partie s'intéresse aux résultats recueillis chez l'ensemble des sujets impliqués alors que

la seconde présente les résultats relevés en fonction du sexe des sujets.

Résultats pour l'ensemble des sujets

Cette recherche révèle des coefficients de consistance interne (coefficient de fidélité Split-half de Spearman-Brown) de l'ordre de 0,968 pour le LAP corrigé selon la méthode AUC et de 0,940 lorsque celui-ci est corrigé selon la méthode INR. Ces indices statistiques illustrent, au moins partiellement, la qualité métrologique de ce test et ils incitent à n'associer aucune restriction majeure à l'interprétation des résultats obtenus à partir de ce prédicteur.

Le tableau 2 présente les résultats moyens et les écarts-types pour chacune des mesures du prédicteur et pour chacun des critères de cette recherche. La consultation de ce tableau révèle des différences en termes de moyennes et d'écarts-types au niveau des différentes mesures au LAP. Certaines de ces différences s'expliquent par le fait que l'on fait appel à deux modes de pondération où l'on obtient un score sur 680 et l'autre sur 800, ce qui implique des moyennes et des écarts-types différents, alors que d'autres différences semblent directement at-

Tableau 2

Moyennes et écarts-types des scores au
Learning Ability Profile et des résultats académiques

	N	Moyenne	Ecart-type
Prédicteurs:			
LAP-680 (AUC)	45	574,42	44,18
LAP-800 (AUC)	45	668,28	51,57
LAP-680 (INR)	45	478,88	71,26
LAP-800 (INR)	45	546,44	83,31
Critères:			
\bar{X} cum. Hiver 1983	45	2,82	0,63
Psy. de la perception I	40	2,47	1,06
Psy. de l'apprentissage	43	2,74	0,98
Histoire de la psy.	42	2,80	0,67
Psychophysiologie I	42	2,66	1,00
Statistiques en psy.	34	2,76	1,23
Psy. dynamique I	38	3,02	0,49
Psy. des groupes I	40	3,07	0,73
Psychophysiologie II	35	3,05	0,76
Psy. génétique I	34	3,00	0,74
Psy. sociale I	37	3,05	0,88

tribuables à la méthode de correction utilisée, soit AUC ou INR. Afin de vérifier l'existence de différences significatives au niveau des moyennes et des écarts-types entre les deux méthodes de correction, ces données ont été soumises au test t et au rapport F (Dayhaw, 1979). Cette analyse statistique a permis de déceler une différence suffisante ($p < 0,01$) dans le rendement moyen ($t=19,01$ et $26,73$ pour le LAP sur 680 et 800 respectivement) et dans la variance ($F=7,66$ et $7,78$) pour admettre la non équivalence des distributions relevées à partir des deux méthodes de correction. Il ressort en effet que les moyennes obtenues à l'aide de la méthode AUC s'avèrent significativement plus élevées que celles obtenues à partir de la méthode INR, cette dernière méthode permettant cependant d'obtenir des résultats avec des écarts-types significativement plus grands que ceux obtenus avec la méthode AUC. La différence entre les distributions se rapportant à chacun des scores au LAP est illustrée à l'appendice B et se concrétise par une distribution plus étendue et symétrique dans le cas de la méthode INR comparativement à la méthode de correction suggérée par l'auteur de ce test (Henning, 1976), à savoir la méthode AUC.

Il convient de souligner que les différences

relevées entre les deux méthodes de correction au niveau des résultats moyens ainsi qu'au niveau des écarts-types constituent un des fondements de l'hypothèse III. Rappelons que cette hypothèse s'intéresse à l'effet de la restriction de la dispersion des scores au prédicteur et en ce sens, elle prédit que l'utilisation de la méthode de correction INR permettra d'obtenir des coefficients de validité plus élevés.

La matrice d'intercorrélations entre les résultats au test selon le mode de pondération (680 ou 800) et la méthode de correction utilisée (AUC ou INR) est exposée au tableau 3. Tous les coefficients de corrélation sont supérieurs à 0,92 et significatifs au seuil de 0,0001. Cette relation qu'on peut qualifier de forte n'est pas surprenante dans le présent contexte étant donné la nature des variables mises en relation. Les résultats de ce tableau suggèrent que le classement relatif d'un sujet par rapport au groupe ne change pas malgré les différents modes de pondération et les différentes méthodes de correction employées pour lui attribuer un score au test psychométrique. De plus, les fortes corrélations (0,984 et 0,989) relevées entre les résultats au LAP pondérés sur 680 et ceux pondérés sur un maximum de 800 points peuvent être

Tableau 3

Intercorrélations entre les résultats au test selon le mode de pondération et la méthode de correction utilisée (N=45)

	LAP-680 (AUC)	LAP-800 (AUC)	LAP-680 (INR)	LAP-800 (INR)
LAP-680 (AUC)	1,000			
LAP-800 (AUC)	0,984	1,000		
LAP-680 (INR)	0,938	0,940	1,000	
LAP-800 (INR)	0,925	0,943	0,989	1,000

considérées comme une forme de support empirique à la méthode de pondération proposée par Henning (1976) afin de compenser pour l'élimination de 12 questions exigeant la connaissance de l'anglais.

D'autre part, le tableau 4 introduit les résultats fondamentaux de la présente recherche puisqu'il indique les coefficients de corrélation relevés entre les scores au test et les résultats scolaires. En premier lieu, ce tableau révèle qu'il n'existe qu'une faible relation entre les résultats au LAP et la moyenne cumulative à l'hiver 1983, cette moyenne représentant les résultats

Tableau 4

Coefficients de corrélation entre les résultats au Learning Ability Profile
et les résultats académiques

Critères	Prédicteurs				N
	LAP-680 (AUC)	LAP-800 (AUC)	LAP-680 (INR)	LAP-800 (INR)	
\bar{X} cum. Hiver 1983	0,165	0,142	0,090	0,087	45
Psy. de la perception I	0,391**	0,368**	0,294*	0,296*	40
Psy. de l'apprentissage	0,095	0,071	0,040	0,022	43
Histoire de la psy.	-0,057	-0,086	-0,113	-0,118	42
Psychophysiologie I	0,380**	0,362**	0,349**	0,335**	42
Statistiques en psy.	0,314*	0,302*	0,306*	0,292*	34
Psy. dynamique I	0,080	0,077	0,005	0,001	38
Psy. des groupes I	-0,020	0,012	-0,036	-0,034	40
Psychophysiologie II	0,203	0,211	0,088	0,097	35
Psy. génétique I	0,189	0,245	0,106	0,150	34
Psy. sociale I	0,100	0,094	0,039	0,051	37

**p < 0,01

*p < 0,05

académiques après deux sessions universitaires. Les coefficients de corrélation rapportés entre ce prédicteur et ce critère varient entre 0,087 et 0,165 selon le mode de pondération ou la méthode de correction utilisée et sont tous non significatifs au seuil de 0,05, ce qui infirme l'hypothèse principale de l'étude (hypothèse I).

Cependant, lorsque le niveau de relation est calculé entre les résultats au prédicteur et ceux obtenus à certains cours individuels constituant la moyenne cumulative, on observe que l'ampleur des coefficients de corrélation varie considérablement d'un cours à l'autre: de coefficients faiblement négatifs on passe à quelques corrélations significativement positives. Ainsi, il semble y avoir une relation substantielle dans le cas des cours: Psychologie de la perception I, Psychophysiologie I et Statistiques en psychologie où l'on obtient respectivement des coefficients de validité prédictive de l'ordre de 0,391 ($p < 0,01$), 0,380 ($p < 0,01$) et 0,314 ($p < 0,05$) entre les notes à ces cours et les résultats au test sur 680 corrigé selon la méthode AUC. En ce qui concerne les autres cours, on remarque une corrélation positive non significative pour les cours Psychophysiologie II ($r=0,203$) et Psychologie génétique I ($r=0,189$) alors qu'une corrélation voisine

de zéro ou faiblement négative est constatée pour les cours: Psychologie de l'apprentissage ($r=0,095$), Psychologie dynamique I ($r=0,080$), Psychologie sociale I ($r=0,100$), Histoire de la psychologie ($r=-0,057$) et Psychologie des groupes I ($r=-0,020$). Il faut noter qu'un profil similaire de corrélations est relevé pour tous ces cours avec les autres modes de pondération et de correction employés au niveau du LAP. Les résultats semblent donc confirmer partiellement l'hypothèse 2 de cette recherche, relative à la validité différentielle, qui prévoyait une meilleure validité prédictive du LAP pour les cours à caractère scientifique plutôt que pour ceux à prédominance verbale.

La consultation de ce tableau permet aussi de constater que les coefficients de validité prédictive obtenus au LAP pondéré sur 680 sont, dans la majorité des cas, légèrement plus élevés que ceux relevés lorsque les résultats à ce test sont rapportés sur un maximum de 800. Il apparaît également que les coefficients de corrélation obtenus lorsque le test psychométrique est corrigé selon la méthode AUC sont tous plus élevés que ceux obtenus lorsque ce même prédicteur est corrigé selon la méthode INR. Le test de signification (t) de la différence entre deux coefficients de corrélation ayant une variable en commun

(Fergusson, 1976) ne révèle toutefois aucune différence significative au seuil de 0,05 entre les différents coefficients relevés selon le mode de pondération utilisé, de même qu'entre les corrélations obtenus selon la méthode de correction employée.

Les différences observées entre l'ampleur des coefficients de validité prédictive selon la méthode de correction utilisée obligent donc à rejeter l'hypothèse III qui voulait que le fait d'obtenir une plus grande variabilité des résultats au prédicteur (méthode INR) permette d'obtenir des coefficients de validité significativement plus élevés.

Par ailleurs, en vue de vérifier l'influence possible d'une restriction des résultats au niveau du critère sur l'ampleur des coefficients de validité prédictive obtenus (hypothèse IV), les coefficients de validité (convertis en valeur z de Fisher) pour chacun des cours ont été mis en relation avec l'écart-type des résultats académiques propre à chacun des cours. De plus, afin d'éprouver l'assertion voulant que lorsque la moyenne des résultats à un cours universitaire est élevée on observe habituellement une faible variabilité au niveau des notes

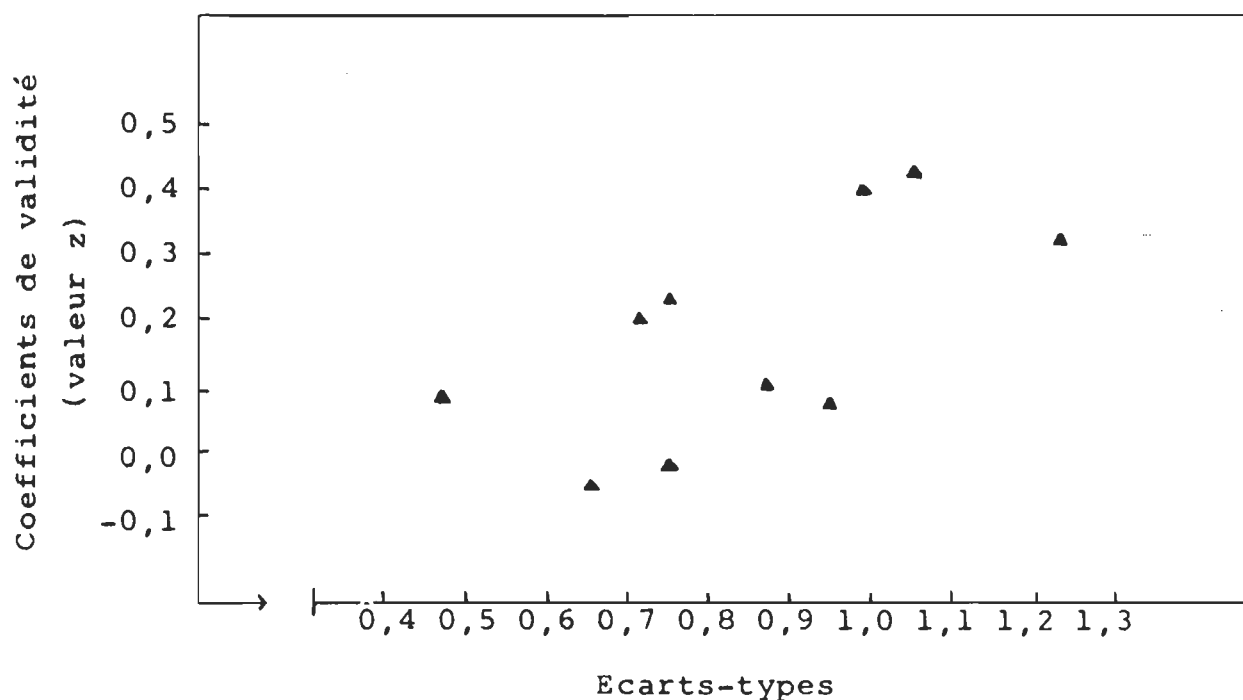


Fig. 1 - Relation entre le coefficient de validité prédictive et l'écart-type des résultats pour un cours universitaire.

attribuées à ces cours (Singleton et Smith, 1978), un coefficient de corrélation a été calculé entre la moyenne des résultats à un cours et l'écart-type de ce cours. Par la suite, les coefficients de validité prédictive étaient mis en relation avec la moyenne des résultats à chacun des cours. Les résultats de cette analyse s'expriment ainsi: un coefficient de corrélation de 0,698 ($p < 0,05$) est relevé

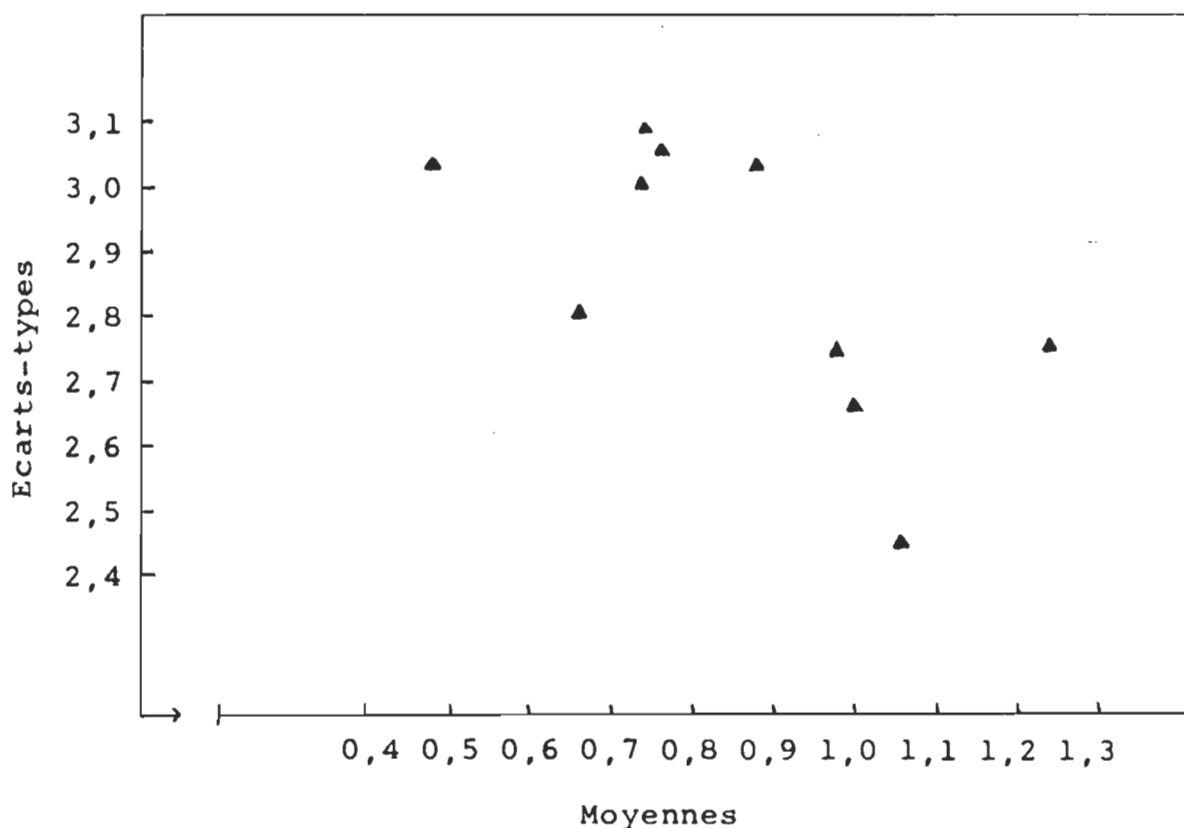


Fig. 2 - Relation entre la moyenne et l'écart-type des résultats pour un cours universitaire.

lorsque les coefficients de validité prédictive sont pairés avec les écarts-types des cours correspondants; une corrélation de $-0,662$ ($p < 0,05$) est observée entre la moyenne des résultats à un cours et l'écart-type à ce cours; finalement, l'analyse des interrelations entre le coefficient de validité à un cours et la moyenne des résultats à ce cours révèle une corrélation de $-0,629$, ce coefficient n'attei-

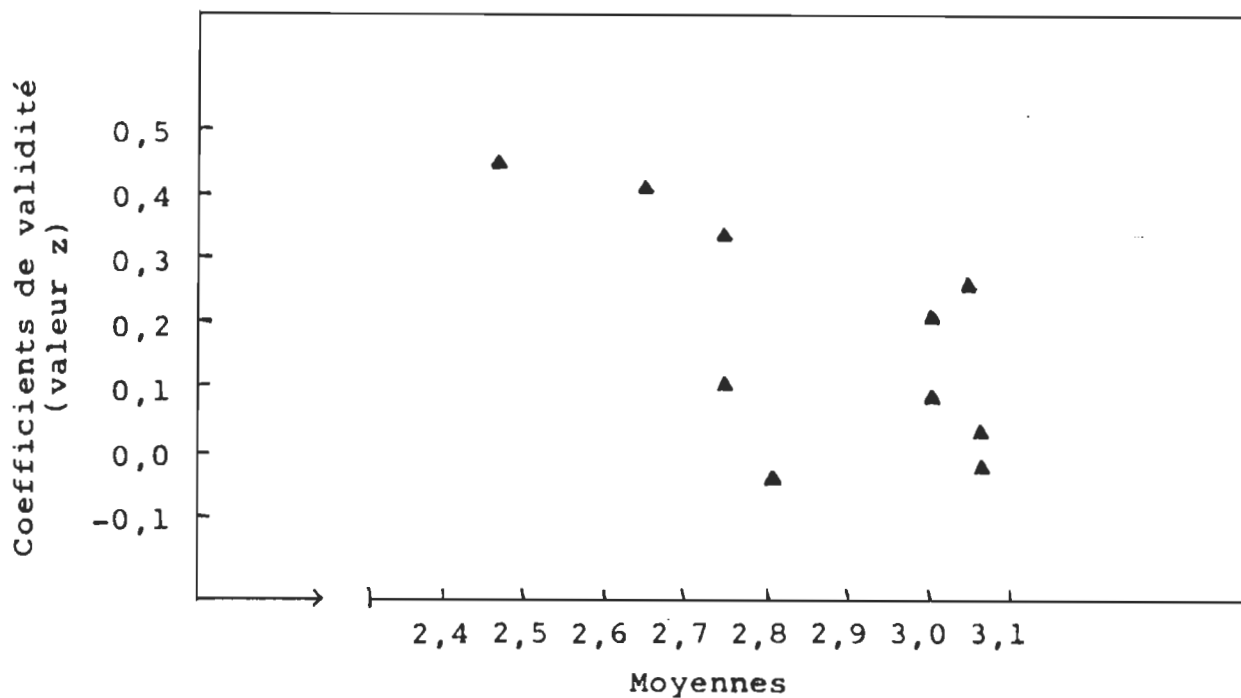


Fig. 3 - Relation entre le coefficient de validité prédictive et la moyenne des résultats pour un cours universitaire.

gnant toutefois pas le seuil de signification de 0,05 étant donné le nombre restreint de paires de mesures impliquées.

L'expression graphique des relations entre ces variables se retrouvent aux figures 1, 2 et 3. La figure 1 présente la relation entre l'écart-type et l'ampleur du

coefficient de validité pour un cours donné. Pour leur part, la figure 2 illustre la relation existante entre la moyenne des résultats à un cours et l'écart-type des résultats au cours correspondant alors que la corrélation entre la moyenne obtenue à un cours et le coefficient de validité obtenu pour ce cours est représentée à la figure 3.

En conformité avec l'hypothèse IV, les résultats de cette analyse statistique révèle l'existence d'une relation significative entre l'ampleur de l'écart-type pour les dix cours considérés et la valeur du coefficient de validité obtenu à ces cours. Ainsi, il semble que les coefficients de validité prédictive les plus élevés sont obtenus dans les cours où l'on relève une grande variabilité au niveau des notes attribuées aux étudiants. Cette analyse corrélationnelle révèle aussi que les cours où l'on observe des résultats moyens élevés sont, dans la majorité des cas, des cours pour lesquels on relève une faible variabilité au niveau des notes attribuées par les professeurs. Il semble que le lien étroit entre la moyenne des résultats à un cours et son écart-type entraîne par le fait même une forte corrélation négative entre la moyenne des résultats à un cours et les coefficients de validité obtenus.

Résultats selon le sexe des sujets

Aux tableaux 5 et 6 sont colligés les résultats moyens ainsi que les écarts-types de chacune des mesures au prédicteur et au critère pour le groupe de sujets féminins et masculins respectivement. La comparaison de ces deux tableaux révèle qu'au niveau des résultats à l'examen psychotechnique, les sujets féminins obtiennent des résultats moyens plus faibles que les sujets masculins alors qu'au niveau des indices de variabilité de ce prédicteur, on remarque que c'est chez le groupe féminin qu'on relève les écarts-types les plus élevés. En ce qui concerne la moyenne cumulative et les résultats aux différents cours, aucune tendance particulière ne peut être relevée par l'observation des moyennes et des écarts-types respectifs entre les deux groupes.

Dans le but de vérifier l'existence de différences significatives au niveau des résultats rapportés aux prédicteurs et aux critères entre les deux groupes de sujets, les résultats moyens ont été comparés à l'aide du rapport t tandis que les différences entre les écarts-types ont été évaluées à partir du rapport F . Les résultats de ces tests de comparaisons indiquent que la valeur des écarts-

Tableau 5

Moyennes et écarts-types des scores au
Learning Ability Profile et des résultats académiques
 pour le groupe de sujets féminins

	N	Moyenne	Ecart-type
Prédicteurs:			
LAP-680 (AUC)	24	565,87	52,55
LAP-800 (AUC)	24	659,12	61,02
LAP-680 (INR)	24	468,33	86,41
LAP-800 (INR)	24	534,58	100,73
Critères:			
\bar{X} cum. Hiver 1983	24	2,83	0,62
Psy. de la perception I	21	2,52	1,17
Psy. de l'apprentissage	23	2,69	0,93
Histoire de la psy.	22	2,91	0,53
Psychophysiologie I	23	2,61	1,03
Statistiques en psy.	18	2,44	1,29
Psy. dynamique I	18	3,22	0,43
Psy. des groupes I	21	2,95	0,74
Psychophysiologie II	17	3,00	0,71
Psy. génétique I	19	3,05	0,70
Psy. sociale I	20	3,15	0,87

Tableau 6

Moyennes et écarts-types des scores au
Learning Ability Profile et des résultats académiques
 pour le groupe de sujets masculins

	N	Moyennes	Ecart-type
Prédicteurs:			
LAP-680 (AUC)	21	584,19	30,51
LAP-800 (AUC)	21	678,76	36,79
LAP-680 (AUC)	21	490,95	47,95
LAP-800 (AUC)	21	560,00	56,92
Critères:			
\bar{X} cum. Hiver 1983	21	2,80	0,66
Psy. de la perception I	19	2,42	0,96
Psy. de l'apprentissage	20	2,80	1,06
Histoire de la psy.	20	2,70	0,80
Psychophysiologie I	19	2,74	0,99
Statistiques en psy.	16	3,12	1,09
Psy. dynamique I	20	2,85	0,49
Psy. des groupes I	19	3,21	0,71
Psychophysiologie II	18	3,11	0,83
Psy. génétique I	15	2,93	0,80
Psy. sociale I	17	2,94	0,90

types au LAP est significativement plus élevée ($p < 0,05$) chez le groupe de sujets féminins, alors que les différences rapportées au niveau des moyennes aux prédicteurs n'atteignent pas le seuil requis pour admettre la non-équivalence de ces indices statistiques entre les deux groupes de sujets. Cette dernière constatation d'équivalence s'applique également en ce qui concerne la moyenne et les écarts-types des résultats académiques.

Les coefficients de validité prédictive relevés lorsque l'on met en relation les scores au LAP avec les résultats académiques pour chacun des deux groupes sont introduits aux tableaux 7 et 8. La consultation du tableau 7, qui se rapporte aux sujets féminins, indique que les coefficients de corrélation varient entre $-0,112$ et $0,636$ selon le critère mis en cause, le mode de pondération ou la méthode de correction employée. Si l'on s'en tient à la relation existante entre les résultats scolaires et le score pondéré sur un maximum de 680 et corrigé selon la méthode AUC, il semble y avoir une forte corrélation entre ces variables dans le cas des cours suivants: Psychologie de la perception I, Psychophysiologie I, Statistiques en psychologie et Psychophysiologie II où l'on remarque des coefficients de l'ordre de $0,636$ ($p < 0,01$), $0,466$ ($p < 0,05$), $0,456$ ($p < 0,05$) et $0,417$ ($p < 0,05$)

Tableau 7

Coefficients de corrélation entre les résultats au Learning Ability Profile
et les résultats académiques pour le groupe de sujets féminins

Critères	Prédicteurs				N
	LAP-680 (AUC)	LAP-800 (AUC)	LAP-680 (INR)	LAP-800 (INR)	
\bar{X} cum. Hiver 1983	0,035	0,290	0,212	0,212	24
Psy. de la perception I	0,636**	0,599*	0,477*	0,442*	21
Psy. de l'apprentissage	0,177	0,178	0,140	0,140	23
Histoire de la psy.	-0,081	-0,074	-0,085	-0,061	22
Psychophysiologie I	0,466*	0,429*	0,393*	0,377*	23
Statistiques en psy.	0,456*	0,456*	0,404*	0,401*	18
Psy. dynamique I	0,037	0,047	0,021	0,036	18
Psy. des groupes I	-0,112	-0,063	-0,109	-0,091	21
Psychophysiologie II	0,417*	0,402*	0,151	0,156	17
Psy. génétique I	0,324	0,362	0,260	0,290	19
Psy. sociale I	0,076	0,059	-0,006	-0,003	20

**p < 0,01

*p < 0,05

respectivement. Des coefficients de validité prédictive pouvant être qualifiés de modérés, soit 0,305 et 0,324 sont rapportés pour la moyenne cumulative à l'hiver 1983 et le cours Psychologie génétique I; ces coefficients n'atteignent toutefois pas le seuil de 0,05 retenu dans cette recherche. Pour ce qui a trait aux autres cours, ce tableau révèle des corrélations positives presque nulles pour les cours Psychologie de l'apprentissage ($r=0,177$), Psychologie dynamique I ($r=0,037$) et Psychologie sociale I ($r=0,076$) alors que des corrélations négatives sont obtenues pour les cours Histoire de la psychologie ($r=-0,081$) et Psychologie des groupes ($r=-0,112$).

Le tableau 8 rapporte les coefficients de validité prédictive obtenus par les sujets masculins du groupe expérimental. L'observation de ce tableau révèle qu'un seul coefficient de corrélation est significatif au seuil de 0,05, ce coefficient étant de 0,488 pour le cours Psychologie dynamique I. Les autres coefficients de validité relevés entre le score au LAP sur 680 (AUC) et les résultats académiques se répartissent ainsi: pour les cours Psychophysiologie I et Psychologie sociale I de faibles corrélations positives sont observées, soit 0,213 et 0,217 respectivement, alors que pour la moyenne cumulative à

Tableau 8

Coefficients de corrélation entre les résultats au Learning Ability Profile
et les résultats académiques pour le groupe de sujets masculins

Critères	Prédicteurs				N
	LAP-680 (AUC)	LAP-800 (AUC)	LAP-680 (INR)	LAP-800 (INR)	
\bar{X} cum. Hiver 1983	-0,51	-0,089	-0,121	-0,127	21
Psy. de la perception I	-0,079	-0,066	-0,020	0,002	19
Psy. de l'apprentissage	-0,060	-0,123	-0,164	-0,211	20
Histoire de la psy.	0,044	-0,048	-0,122	-0,166	20
Psychophysiologie I	0,213	0,232	0,268	0,257	19
Statistiques en psy.	-0,128	-0,168	-0,022	-0,076	16
Psy. dynamique I	0,488*	0,446*	0,260	0,230	20
Psy. des groupes I	0,005	0,022	-0,008	-0,039	19
Psychophysiologie II	-0,012	0,208	0,009	0,024	18
Psy. génétique I	0,056	0,153	-0,101	-0,017	15
Psy. sociale I	0,217	0,217	0,173	0,203	17

* $p < 0,05$

l'hiver 1983 et les autres cours, les coefficients de corrélation ont une valeur positive ou négative presque nulle.

Lorsque l'on compare les données rapportées par ces deux tableaux, un seul point en commun semble ressortir: les différentes méthodes de pondération et de correction utilisées pour déterminer le score total à l'examen psychométrique n'ont qu'un très faible effet sur l'ampleur des coefficients de validité obtenus. Dans la majorité des cas, les coefficients de corrélation sont légèrement plus élevés lorsque le test est corrigé selon la méthode AUC et qu'il est pondéré sur un maximum de 680 points.

D'autre part, il semble important d'attirer l'attention sur le fait que pour un même critère du rendement scolaire, les résultats rapportés en termes de coefficients de corrélation pour les sujets féminins sont sensiblement différents de ceux obtenus chez les sujets masculins. Afin d'analyser ces différences, les coefficients de validité prédictive obtenus par l'ensemble des sujets ainsi que ceux relevés plus spécifiquement pour les sujets féminins et masculins du groupe expérimental sont regroupés au tableau 9 mais cette fois en ne retenant que le LAP-680 corrigé selon la méthode AUC comme prédicteur.

A la lecture de ce tableau, on constate que les cours universitaires où le groupe de sujets féminins obtient les corrélations les plus élevées tendent à correspondre aux cours pour lesquels les sujets masculins obtiennent les coefficients les plus faibles, et vice versa. Par exemple, pour les cours Psychologie de la perception I, Statistiques en psychologie et Psychophysologie II on observe pour le groupe de sujets féminins des coefficients de corrélation de 0,636, 0,456 et 0,417 respectivement, alors que pour le groupe masculin on rapporte pour les mêmes cours des coefficients de l'ordre de -0,079, -0,128 et -0,012. De plus, comparativement au groupe de sujets masculins, la valeur des coefficients de validité relevés pour le groupe féminin à chacun des critères se rapproche beaucoup plus de ceux rapportés pour l'ensemble des sujets.

Ce tableau permet aussi d'observer que les sujets féminins du groupe expérimental obtiennent des coefficients de validité plus élevés que les sujets masculins et ce, pour la moyenne cumulative ainsi que six cours sur une possibilité de dix. Les différences relevées entre l'ampleur des coefficients de validité entre les deux groupes ont été analysées à partir de l'erreur-type d'une différence entre deux coefficients z de Fisher (Dayhaw, 1979) et ce

Tableau 9

Coefficients de corrélation obtenus entre le score au LAP-680 (AUC) et le rendement académique pour les sujets féminins et masculins ainsi que pour l'ensemble du groupe expérimental

	Sujets féminins (N=24)	Sujets masculins (N=21)	Groupe total (N=45)
\bar{X} cum. Hiver 1983	0,305	-0,051	0,165
Psy. de la perception I	0,636**	-0,079	0,391**
Psy. de l'apprentissage	0,177	-0,060	0,095
Histoire de la psy.	-0,081	0,044	-0,057
Psychophysiologie I	0,466*	0,213	0,380**
Statistiques en psy.	0,456*	-0,128	0,314*
Psy. dynamique I	0,037	0,488*	0,080
Psy. des groupes I	-0,112	0,005	-0,020
Psychophysiologie II	0,417*	-0,012	0,203
Psy. génétique I	0,324	0,056	0,189
Psy. sociale I	0,076	0,217	0,100

**p < 0,01

*p < 0,05

test statistique n'a permis de déceler une différence significative ($p < 0,05$) que pour le cours Psychologie de la perception I où l'on observe un coefficient de 0,636 pour les sujets féminins alors qu'un coefficient de validité de -0,079 est relevé pour les sujets masculins.

De telles différences au niveau de l'ampleur des

corrélations remarquées entre les deux groupes d'étudiants ne confirme donc que partiellement l'hypothèse V puisque celle-ci prévoyait que les coefficients de validité seraient plus élevés pour le groupe de sujets féminins et cela pour l'ensemble des cours considérés. Il semble toutefois évident que le nombre restreint de sujets impliqués a empêché de confirmer statistiquement une tendance plus nette qui se dégage à la lecture du tableau 9.

Interprétation des résultats

Pour terminer ce chapitre, chacune des cinq hypothèses sous-jacentes à la présente recherche est examinée et discutée à la lumière des résultats précédemment exposés.

La première hypothèse prévoyait qu'il existerait une relation significative entre les scores au Learning Ability Profile (LAP) et la moyenne cumulative des résultats scolaires de deux sessions universitaires. Cette hypothèse se réfère à l'objectif fondamental de cette recherche qui était d'évaluer la validité prédictive de ce test en contexte académique.

Dans l'ensemble, les présents résultats ne confirment pas l'hypothèse principale de cette recherche en indiquant l'absence de relation significative entre les résultats au LAP et la moyenne cumulative de première année universitaire chez un groupe d'étudiants en psychologie. Le rendement à ce test d'apprentissage n'est corrélié de façon significative qu'aux résultats d'un nombre limité de cours suivis par les sujets. Cependant, un bon nombre d'indices théoriques et techniques permettent d'expliquer, au moins partiellement, la nature des résultats obtenus.

Dans la perspective d'expliquer la faiblesse relative des coefficients de validité relevés entre les scores au LAP et la moyenne cumulative, il semble important de mentionner en premier lieu que le schéma expérimental ne permettait pas ici d'envisager autre chose que des coefficients de corrélation relativement faibles. Il a déjà été souligné au premier chapitre de cette étude que le plafond de 0,40 a rarement été dépassé dans les procédures de validation en contexte de prédiction du rendement d'une première année universitaire et cela même pour des tests conçus spécifiquement à cette fin tels que le Scholastic Aptitude Test (SAT) ou les Tests d'Aptitudes aux

Etudes Universitaires (TAEU). Une telle procédure de validation consiste en effet à établir une corrélation entre un prédicteur et un critère, ce dernier concept constituant un indice du rendement humain. La formule suivante résume les principaux paramètres: "Rendement = f [(aptitudes + connaissances) motivation]".

Notre critère, la moyenne cumulative de dix cours universitaires, représente donc un concept multidéterminé par des composantes motivationnelles et cognitives. Or le LAP ne mesure qu'une seule forme de composantes cognitives, à savoir la capacité d'apprentissage. On ne peut donc s'attendre à obtenir un coefficient démesurément élevé puisque'une large proportion de la variance associée au critère est expliquée par des variables indépendantes qui échappent à la mesure dans le cadre de cette recherche (Fabi, 1983).

De plus, il convient de rappeler que dans un contexte de prédiction du rendement scolaire, les coefficients de validité prédictive semblent varier considérablement d'une recherche à l'autre et cela, même pour des tests psychométriques utilisés à très grande échelle comme le SAT, le ACT et le GRE. A titre d'exemple, un examen

de la documentation effectué en 1968 par l'Educational Testing Service (Lanholm et al., 1968: voir Littlepage et al., 1978) révèle que les coefficients de corrélation entre d'une part, le GRE-V, le GRE-Q et le GRE-P et d'autre part, les résultats des étudiants inscrits aux études supérieures en psychologie, varient d'un coefficient de corrélation positif moyen (0,35, 0,45 et 0,35 respectivement) à un coefficient de corrélation négatif (-0,28, -0,24 et -0,35 respectivement). Il ne faut donc pas se surprendre outre mesure si nous notons ici des écarts entre les résultats de cette recherche et ceux obtenus par Fabi (1983) sur la validité du LAP. En ce sens, même si les présents résultats ne révèlent qu'une corrélation de 0,165 ($p > 0,05$) entre le rendement au LAP et la moyenne cumulative, ceux-ci ne nous permettent pas de porter un jugement trop catégorique sur la validité de pronostic de ce test.

Toujours dans la perspective d'expliquer la faiblesse relative des coefficients de validité relevés entre les scores au LAP et la moyenne cumulative, il semble également opportun de s'intéresser à la nature des items contenus dans ce test psychométrique. Rappelons que ce prédicteur est presque essentiellement composé

d'items qui font intervenir des aptitudes dites non-verbales. De ce fait, il convient de s'interroger sur la possibilité que celui-ci recouvre toutes les habiletés impliquées dans la réussite académique.

Le choix de n'inclure que ce genre d'items est motivé par le fait qu'en élaborant ce test d'apprentissage, l'auteur (Henning, 1976) avait comme principale préoccupation de développer une épreuve qui serait libre de toutes influences culturelles en rapport avec la race, le groupe ethnique, le niveau d'éducation ou le niveau socio-économique des candidats. Ce type de test qu'on nomme culture-fair test s'inspire de la théorie de l'intelligence de Cattell (1967) qui veut que l'intelligence se divise en deux facteurs généraux: l'intelligence fluide qui est fortement reliée à l'hérédité et l'intelligence cristallisée qui est reliée à l'influence de l'environnement. Dans le cas du LAP, les aptitudes évaluées tendent à correspondre davantage à un de ces facteurs, soit l'intelligence fluide. Une des méthodes pour mesurer ce type d'intelligence, et celle qui a été utilisée dans une large mesure pour ce test, consiste à avoir recours à des items qui prennent la forme d'un matériel perceptuel plutôt qu'à des questions qui font appel à des aptitudes verbales (Anastasi,

1982; Cronbach, 1972). Or les recherches sur l'intelligence fluide démontrent que ce facteur ne constitue pas un prédicteur très puissant pour faire le pronostic du rendement scolaire (Cattell et Butcher, 1968; Cattell et al., 1972; Eno, 1978).

Pour leur part, les tests traditionnels qui sont réputés être les meilleurs prédicteurs de la moyenne cumulative à l'université mesurent à la fois l'intelligence fluide et cristallisée dans des proportions diverses. Un fait encore plus important, les recherches en ce domaine démontrent que ce sont les sous-tests verbaux qui permettent de faire le meilleur pronostic du succès scolaire. Par exemple, nombre d'auteurs (Conry et Plant, 1965; Quereshi et Miller, 1970; Zung et Gianturco, 1968) se sont intéressés à la validité prédictive du Wechsler Adult Intelligence Scale (WAIS) en contexte académique et ils ont établi que ce sont les sous-tests Vocabulary, Information et Similarities qui sont les meilleurs prédicteurs de la moyenne cumulative au collège et à l'université. Mentionnons que ces trois sous-tests font tous partie de la section verbale de ce test d'intelligence et qu'ils font appel dans une large mesure à des connaissances acquises.

La supériorité du pouvoir prédictif des sous-tests verbaux pour prédire un critère multidimensionnel tel que la moyenne cumulative de première année à l'université se vérifie également pour certains examens psychométriques construits spécifiquement à cette fin. Ainsi, une synthèse de plus de 700 études sur la validité du SAT effectuée par Campos (1977: voir Slack et Porter, 1980) révèle un coefficient de validité prédictive moyen de 0,40 pour le SAT-V et de 0,35 pour le SAT-M.

Lorsque l'on compare le contenu des principaux tests prévus aux fins de sélection des étudiants avec celui du LAP, un autre fait important semble se dégager: les items rencontrés dans le SAT, le GRE et le ACT se rapprochent beaucoup plus des tâches réelles effectuées dans le cadre des études universitaires. En ce qui concerne le ACT, cette relation entre le contenu du test et les diverses matières académiques se comprend aisément puisque celui-ci étant par définition un test d'achievement. Par contre, le SAT et le GRE sont des tests d'aptitudes au même titre que le LAP et contrairement à celui-ci, ils contiennent de nombreuses questions qui ont un rapport direct avec les tâches requises dans certaines matières scolaires. Cette relation de contenu s'applique

particulièrement aux sous-tests Reading comprehension et Mathematics de ces deux tests d'aptitudes scolaires.

Selon Bégin (1980), la notion d'"aptitude" doit faire référence à un facteur inné et stable qui est modifiable en partie par les conditions de l'environnement. Si l'on adopte cette définition, le LAP se rapprocherait beaucoup plus d'une mesure d'aptitude que ne le sont le GRE et le SAT. Bien que ce manque de relation intrinsèque entre le contenu de l'épreuve psychométrique et la tâche des étudiants est souhaitable voir même nécessaire pour un test d'aptitude comme le LAP qui se veut libre d'influences culturelles, il n'en demeure pas moins que cela peut avoir une influence négative sur le pronostic du rendement académique. Cela se vérifie dans la mesure où, actuellement, ce sont les résultats scolaires antérieurs et les résultats à des tests de rendement scolaire qui demeurent les meilleurs moyens de prévision du succès académique ultérieur (Boucher, 1982).

Cette discussion sur le contenu du LAP nous introduit d'ailleurs aux limites inhérentes à ce que certains appellent la psychométrie traditionnelle, limites qui sont de plus en plus fréquemment énoncées dans la

documentation (Bégin, 1980; Tryon, 1979; Tyler, 1978: voir Boucher, 1982). Ces auteurs expliquent l'incapacité dans laquelle se trouvent les chercheurs d'améliorer de façon notable la validité de pronostic des instruments de mesure en invoquant le fait que la psychométrie, dans le contexte de la psychologie appliquée, s'est orientée progressivement vers l'identification des caractéristiques, des construits, des traits susceptibles d'expliquer le comportement, le rendement des individus à des tâches de nature scolaire et que, ce faisant, elle a imposé comme allant de soi une conception et une interprétation des différences individuelles. Ceux-ci suggèrent, à titre d'alternative, l'approche des compétences qui préconise un retour à la pratique de l'analyse de la tâche pour déterminer les habiletés requises à l'exécution d'une tâche donnée, puis à développer un instrument ayant une validité de contenu par rapport à la tâche considérée.

L'application de ce modèle au problème du pronostic du rendement scolaire permet de constater que l'une des principales tâches des étudiants consiste à acquérir des connaissances et à les comprendre. En conséquence, un test standardisé servant à prédire le rendement académique de première année universitaire devrait donc soumettre les

étudiants à une tâche d'apprentissage du genre de celle qu'ils auront à effectuer durant leurs études et mesurer l'acquisition et la compréhension de connaissances (Boucher, 1982).

Il semble toutefois important de souligner que les qualités métrologiques et l'originalité du LAP ne résident pas spécialement dans la nature des items qu'il contient mais plutôt dans la feuille réponses qui l'accompagne. C'est par l'entremise de la technique carbon erasure procedure qui permet de donner un feedback concernant l'exactitude de la réponse que cette épreuve psychométrique devient une mesure de la capacité d'apprentissage. Concrètement, une telle technique permet à l'évalué d'apprendre certains principes de raisonnement à mesure qu'il effectue le test, la situation d'évaluation se transformant en une véritable situation d'apprentissage (Henning, 1976). Outre l'originalité de son mode de réponse, le LAP présente une excellente consistance interne qui témoigne de l'homogénéité des items de ce test.

Ces commentaires sur le contenu du LAP nous amènent à discuter les résultats obtenus en rapport avec l'hypothèse II de cette recherche. Celle-ci stipulait

que l'ampleur des coefficients de validité prédictive obtenus à partir des cours universitaires à caractère scientifique sera plus élevés que ceux obtenus pour les cours à prédominance verbale.

Cette hypothèse fait référence à la validité différentielle du LAP. La formulation de cette hypothèse n'était soutenue par aucune évidence empirique concernant la possibilité que ce test puisse prédire avec plus d'acuité certains types de cours. Rappelons aussi que les résultats obtenus à partir de tests comme le SAT, le GRE et le ACT sont équivoques quant aux possibilités de prédire d'une façon différentielle la réussite dans des cours spécifiques. L'hypothèse II ne reposait donc sur le fait que, ce prédicteur étant en grande partie composé d'items qui font appel à des aptitudes dites non-verbales, on devait s'attendre à ce que les scores au LAP soient plus fortement corrélés avec les résultats aux cours à caractère scientifique comparativement aux cours à prédominance verbale. En l'absence de définitions concrètes de ces deux termes qui sont fréquemment retrouvés dans la documentation abordant cette problématique (Carter, 1972; Goldman et Slaughter, 1976; Passons, 1967), cette étude a proposé de définir un cours à caractère scientifique

comme étant un cours qui se caractérise par la présentation et l'évaluation de données objectives et de connaissances exactes alors qu'un cours à prédominance verbale a été défini comme étant une matière scolaire où les connaissances enseignées sont sujettes à discussion, à l'interprétation et où il est souvent possible de donner plusieurs réponses à un même problème.

Les résultats en rapport avec cette hypothèse révèlent, par ordre d'importance, les coefficients de validité suivants: Psychologie de la perception I ($r=0,391$, $p<0,01$), Psychophysiologie I ($r=0,380$, $p<0,01$), Statistiques en psychologie ($r=0,314$, $p<0,05$), Psychophysiologie II ($r=0,203$), Psychologie génétique I ($r=0,189$), Psychologie sociale I ($r=0,100$), Psychologie de l'apprentissage ($r=0,095$), Psychologie dynamique I ($r=0,080$), Psychologie des groupes I ($r=-0,020$) et Histoire de la psychologie ($r=-0,057$).

Bien qu'il soit difficile de classifier tous les cours selon la dichotomie proposée à l'hypothèse II, il semble que le profil de corrélation observé tend à confirmer cette hypothèse. Ainsi on remarque que des cours à caractère plutôt scientifique tels que le cours Psy-

chologie de la perception I, Psychophysiologie I et II et Statistiques en psychologie obtiennent des corrélations beaucoup plus élevées que celles relevées pour des cours à prédominance verbale comme le cours Psychologie dynamique I, Psychologie des groupes I et d'Histoire de la psychologie.

Toutefois, certaines réserves doivent être énoncées en relation avec cette hypothèse. D'une part, le choix de classifier un cours comme étant une matière scientifique ou verbale ne repose sur aucune analyse formelle du contenu des cours considérés. D'autre part, les cours pour lesquels on relève les plus forts coefficients de corrélation tendent à correspondre aux cours où l'on observe une grande variabilité dans les notes attribuées par les enseignants (voir hypothèse IV). Il est donc difficile d'établir l'importance relative de ces deux facteurs. Pour ces raisons, il semble plus prudent ici d'avancer que les résultats confirment partiellement l'hypothèse II de cette recherche.

Pour leur part, les hypothèses III, IV et V visaient à évaluer l'effet de certaines variables sur l'ampleur des coefficients de validité prédictive obtenus

à partir du LAP. La première variable étudiée se rapporte à l'influence de la grandeur de l'écart-type des scores au test psychométrique sur l'ampleur des coefficients de validité prédictive. L'étude de cette variable était motivée par le fait qu'un écart-type relativement faible au niveau des résultats au LAP avait été observé dans l'étude de Fabi (1983) et que cet auteur avait soupçonné ce facteur d'avoir eu un effet atténuateur sur l'ampleur des coefficients de validité observés.

Afin d'analyser l'effet possible d'une restriction de la dispersion des résultats au prédicteur sur l'ampleur des coefficients de validité, cette recherche a fait intervenir deux méthodes de correction pour le test. La première méthode de correction, et celle proposée par l'auteur de cette échelle (Henning, 1976), exige que l'évalué continue à répondre à chacune des questions jusqu'à l'obtention de la bonne réponse, un score étant établi selon le nombre d'essais requis. Cette méthode nommée answer-until-correct (AUC) permet donc aux sujets d'obtenir un score de 9, 7, 6, 4, 3, 1 et 0 sur 10 même s'ils n'ont pas répondu correctement au premier essai, ce qui peut expliquer en partie la faible dispersion des résultats au prédicteur. La deuxième méthode de correction

employée, inferred number right score (INR), fait intervenir le même processus pour répondre à chacun des items mais cette fois, on ne considère que le premier choix de réponse du sujet et un score de 10 ou 0 est alloué selon l'exactitude de sa réponse. En prenant pour acquis que cette deuxième méthode de correction permettra éventuellement d'obtenir un écart-type plus élevé au niveau des résultats au LAP, l'hypothèse suivante avait été formulée: les coefficients de validité prédictive obtenus au LAP à partir de la méthode de correction INR seront significativement plus élevés que ceux relevés à partir de la méthode AUC.

Les résultats relatifs à cette hypothèse révèlent que même si l'utilisation de la méthode INR a permis d'augmenter significativement l'ampleur de l'écart-type des résultats au LAP tout en maintenant le même processus qui permet à l'évalué d'apprendre certains principes de raisonnement à mesure qu'il effectue l'épreuve psychométrique, les coefficients de validité prédictive n'ont pas été significativement plus élevés que ceux obtenus avec la méthode AUC. Les coefficients de validité relevés pour chacune des deux méthodes étant dans la plupart des cas de valeur équivalente, l'hypothèse III de cette recherche est donc infirmée.

Cependant, dans la mesure où l'on peut comparer l'écart-type relevé au LAP avec ceux rapportés pour des tests comme le SAT et le LSAT qui sont eux aussi pondérés sur un maximum de 800 points, l'effet modérateur d'une telle variable sur l'ampleur des coefficients de validité prédictive peut être invoqué. Pour ces deux tests psychométriques conçus aux fins de prédire le succès scolaire, l'écart-type doit être de 100 pour que les auteurs éliminent la possibilité de l'effet de la restriction des scores au prédicteur (Linn, 1982; Schmidt et Hunter, 1977). Cette valeur est loin d'être atteinte pour le LAP dans le cadre de cette étude.

La deuxième variable étudiée en rapport avec les facteurs explicatifs susceptibles d'avoir une influence sur l'ampleur des coefficients de validité concerne la fidélité et la validité du critère utilisé. Concrètement, elle a trait au caractère non discriminant de certaines évaluations du corps professoral où l'on observe dans certains cas une forte propension à attribuer des cotes A et B comme résultat scolaire. Selon certains auteurs (Dole et Baggaley, 1979; Hartnett et Willingham, 1980; Merenda et Reilly, 1971; Michael et al., 1983; Thacker et Williams, 1974), une telle pratique d'évalua-

tion qui se caractérise par une faible variabilité des résultats au critère ne peut qu'avoir une influence négative sur l'ampleur des coefficients de validité prédictive. Considérant l'importance de ce facteur, l'hypothèse suivante avait été formulée: il existera une relation significative entre l'ampleur de l'écart-type des résultats à un cours universitaire et la grandeur du coefficient de validité prédictive obtenu pour ce cours.

Les résultats obtenus dans cette recherche confirment l'hypothèse IV: un coefficient de corrélation de 0,698 ($p < 0,05$) est relevé lorsque l'écart-type des résultats à un cours est mis en relation avec le coefficient de validité prédictive obtenu pour ce cours. Dans la majorité des cas, les coefficients de validité les plus élevés ont été obtenus dans les cours où prévaut une pratique de grande dispersion des résultats; cette pratique se concrétisant sur le plan statistique par des résultats moyens relativement faibles avec un écart-type assez important.

Dans le cadre de cette étude, une faible variabilité au niveau des résultats scolaires a été remarquée dans plusieurs cours universitaires. Pour certains cours,

on relève que plus de 70% des notes attribuées se regroupent dans la seule cote B (voir appendice B). Le manque de fidélité et de validité de certains critères nous apparaît donc comme un des principaux facteurs pouvant expliquer la faiblesse des coefficients de corrélation obtenus entre les scores au LAP et les résultats à certains cours universitaires.

Enfin, la dernière variable retenue pour fin d'analyse concerne les différences relevées dans la documentation scientifique entre l'ampleur des coefficients de validité obtenus chez les sujets féminins comparativement à ceux obtenus par des sujets masculins (Astin, 1971; Jones, 1970; Khan, 1973; Lavin, 1965). Les différences s'établissant en faveur des sujets féminins, l'hypothèse suivante avait donc été formulée: les coefficients de validité prédictive obtenus par les sujets féminins seront plus élevés que ceux relevés pour les sujets masculins et ce, pour l'ensemble des cours considérés.

Si l'on s'en tient aux résultats empiriques obtenus dans cette recherche, l'hypothèse V semble partiellement confirmée. Les sujets féminins du groupe expérimental obtiennent des coefficients de validité pré-

dictive plus élevés que les sujets masculins pour la moyenne cumulative ainsi que pour six cours sur une possibilité de dix.

Toutefois, lorsque l'on analyse les différences observées dans cette recherche en regard des raisons communément invoquées dans la documentation, de telles différences apparaissent comme étant fortuites. Les auteurs (Astin, 1971; Khan, 1973) expliquent les différences entre l'ampleur des coefficients de validité relevés par ces deux groupes par le fait que les sujets féminins obtiennent des résultats moyens plus élevés au test psychométrique ainsi qu'au niveau des résultats académiques. Dans cette recherche, les comparaisons effectuées entre les deux groupes ne révèlent aucune différence significative au niveau des résultats moyens au prédicteur et au critère.

Il nous semble que les différences observées entre les sexes dans cette étude peuvent être attribuables au petit nombre de sujets impliqués dans les deux groupes. Comme le soulignent Callender et Osburn (1980), un échantillon trop restreint permet l'influence de certains facteurs non contrôlés sur la variable dépendante. Selon la direction que prend cette influence, impondérable, la

valeur de la relation entre la variable indépendante et la variable dépendante peut être tantôt sous-estimée tantôt sur-estimée.

Résumé et conclusion

L'objectif majeur de cette recherche consistait à évaluer la validité prédictive d'un instrument récemment élaboré (Henning, 1976), le Learning Ability Profile (LAP), qui permet d'obtenir une mesure de la capacité d'apprentissage. En second lieu et à titre purement exploratoire, cette étude s'intéressait à la validité différentielle de ce test ainsi qu'à l'influence de certains facteurs susceptibles de faire varier l'ampleur des coefficients de validité.

Ces objectifs ont amené la formulation de cinq hypothèses. Dans le cas des deux premières hypothèses relatives à la validité prédictive (hypothèse I) et différentielle (hypothèse II) du LAP, il s'agissait de calculer des coefficients de corrélation (Pearson) afin de quantifier l'ampleur de la co-variation entre les scores au test et les résultats académiques exprimés sous forme de moyenne cumulative ou de résultats à des cours universitaires. Les résultats de cette analyse révèlent que les coefficients de corrélation obtenus entre les résultats au LAP et la moyenne cumulative n'atteignent pas le seuil de signification requis ($p < 0,05$), ce qui

infirme l'hypothèse I de cette recherche.

Par contre, lorsque les scores à ce test sont mis en relation avec les résultats à des cours individuels, on observe des corrélations significatives pour trois cours sur une possibilité de dix. Les cours universitaires les mieux prédits tendent à correspondre à des matières scolaires à caractère plus scientifique que verbal. Ces résultats semblent donc confirmer l'hypothèse II qui postulait que le LAP prédirait avec plus d'acuité les cours à caractère scientifique étant donné que ce test est en grande partie composé d'items qui font appel à des aptitudes non-verbales ou quantitatives. Cette dernière constatation pourrait, dans une certaine mesure, expliquer les faibles corrélations obtenues entre ce test et un critère multidéterminé comme la moyenne cumulative.

Les trois dernières hypothèses concernaient l'étude de certaines variables qui sont reconnues dans la documentation scientifique comme étant des facteurs qui font varier l'ampleur d'un coefficient de corrélation entre un prédicteur et un critère. La troisième hypothèse s'intéressait plus particulièrement à l'effet de l'étendue des scores au test sur l'ampleur des coeffi-

cients de validité. A cette fin, deux méthodes de correction pour le LAP ont été utilisées et des comparaisons entre les résultats obtenus à partir de ces deux méthodes ont été effectuées. Les résultats relatifs à cette hypothèse révèlent que même si l'une des méthodes de correction (INR) a permis d'obtenir une plus grande variabilité au niveau des scores au test, les coefficients de validité obtenus à partir de cette méthode n'ont pas été significativement plus élevés que ceux relevés pour l'autre méthode de correction (AUC) où l'on avait remarqué un écart-type relativement faible. En infirmant l'hypothèse III, ces résultats nous empêchent d'invoquer le phénomène de la restriction de la dispersion des scores au prédicteur comme explication possible des faibles coefficients de validité obtenus pour certains critères.

En contrepartie, la quatrième hypothèse de cette recherche voulait vérifier l'influence d'une faible variabilité des résultats au critère sur l'ampleur des coefficients de validité prédictive. Pour ce faire, des coefficients de corrélation ont été calculés entre les écarts-types des résultats aux différents cours et les coefficients de validité obtenus pour ces cours. Cette hypothèse a été confirmée, une corrélation élevée et significative

($p < 0,05$) étant relevée entre ces deux variables. Ces résultats nous incitent à considérer le faible pouvoir discriminant de certaines évaluations du corps professoral comme l'une des principales variables pouvant expliquer la faiblesse relative de certains coefficients de validité prédictive obtenus.

La dernière hypothèse se rapportait aux différences observées dans la documentation entre l'ampleur des coefficients de validité prédictive relevés chez les sujets féminins comparativement à ceux obtenus pour les sujets masculins. L'hypothèse V semble partiellement confirmée, les sujets féminins obtenant des coefficients de validité plus élevés que les sujets masculins pour la moyenne cumulative ainsi que pour six cours universitaires sur une possibilité de dix. Les différences observées nous semble toutefois attribuables au petit nombre de sujets impliqués dans les deux groupes.

A titre de conclusion, les résultats obtenus dans cette recherche ne constituent qu'un faible support concernant la validité du LAP en tant que prédicteur du rendement scolaire en milieu universitaire. En considérant le fait que des corrélations significatives ont déjà

été observées entre ce test et la moyenne cumulative (Fabi, 1983) et que l'ampleur des coefficients de validité prédictive relevés pour un test peut varier considérablement d'une recherche à l'autre (Munday, 1970), il nous semble que des études de validation supplémentaires s'imposent avant d'en arriver à des conclusions définitives sur la pertinence de son utilisation en contextes académique ou organisationnel.

Dans l'éventualité où une telle démarche de validation serait répétée, il serait souhaitable que les sujets du groupe expérimental proviennent de facultés où prévaut une pratique de grande dispersion des résultats et où la philosophie d'évaluation laisse le moins de place possible à la subjectivité des évaluateurs. Ces précautions supplémentaires permettraient d'obtenir des résultats plus concluants quant à l'utilité de ce test psychométrique dans une perspective de sélection ou d'orientation scolaire.

Il s'avérerait également opportun de vérifier, comme le suggèrent nos résultats, si les coefficients de validité obtenus à partir du LAP ne seraient pas sensiblement augmentés dans d'autres contextes universitaires où les disciplines enseignées se rapprochent davantage du pôle

scientifique et font moins appel aux aptitudes verbales. Dans cette perspective, on pourrait envisager des études de validation dans des facultés comme la chimie, la physique, la biologie ou l'ingénierie.

Appendice A

Examen psychotechnique



Form B

Version Française



LAP

Learning Ability Profile

Margherita M. Henning Ph.D

Consigne

Ceci n'est pas un examen ordinaire. C'est un exercice pour évaluer votre capacité d'apprentissage. Les 68 problèmes contenus dans ce livret suivent une progression sensée. Il est à noter qu'on ne répond pas aux questions: 4, 18, 22, 52, 60, 63, 67, 69, 72, 74, 76 et 78 qui demandent la connaissance de la langue anglaise.

Vous avez tous un livret de test, une feuille de réponses, un grattoir et un questionnaire. Vous avez aussi besoin d'un stylo à bille pour remplir le questionnaire.

Si vous regardez l'échantillon 0 en haut de la feuille de réponses, vous verrez 4 carrés foncés. Assurez-vous que votre feuille de réponses soit bien appuyée sur une surface dure. En vous servant du grattoir qui vous a été fourni, vous pouvez facilement gratter les carrés foncés de la façon suivante.



Les numéros, -6 -1 10 -3

apparaîtront. Cependant dans l'exercice, les numéros ne suivent pas le même ordre.

Les nombres représentent la QUALITE de vos choix. Le MEILLEUR choix est le nombre 10. Le choix NUMERO DEUX est le nombre -1. Le PROCHAIN choix est le nombre -3, et le choix le PIRE pour n'importe quel problème est le nombre -6. Dans votre intérêt, même si votre premier choix n'est pas le carré identifié par le numéro 10, découvrez le nombre suffisamment sous chaque carré choisi dans chaque réponse afin de voir quel est ce nombre.

Regardez votre feuille de réponses promptement pour vous assurez que les 68 groupes de carrés numérotés soient foncés et que vous ne pouvez voir les numéros en dessous.

Le but de cet exercice est d'utiliser le moins d'essais possibles avant de découvrir le MEILLEUR choix qui sera le nombre 10 dans chacun des 68 problèmes. Mais vous DEVEZ CONTINUER à gratter les carrés de chaque problèmes jusqu'à ce que vous trouviez le nombre 10, et SEULEMENT ALORS pourrez-vous sauter au prochain problème.

Voici quelques-uns des signes qui sont utilisés dans l'exercice:

LE SIGNE + VEUT DIRE "ADDITIONNEZ"

LE SIGNE - VEUT DIRE "ENLEVER" ou "SOUSTRAIER"

LE SIGNE X VEUT DIRE "MULTIPLIEZ"

LE SIGNE = VEUT DIRE "LE MEME" ou "EGALE"

LE SIGNE : VEUT DIRE "EST A"

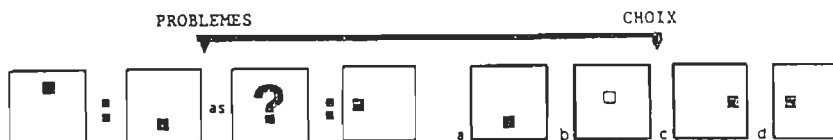
LE SIGNE ? VEUT DIRE "UN DES CHOIX VA A CET ENDROIT"

Par exemple, un NAVIRE est à l'EAU ce qu'un AVION est à l'AIR. Un NAVIRE: l'EAU ce qu'un avion: l'AIR.

Soyez sûrs de vous souvenir des divers signes. Si nécessaire, vous devrez revoir cette page tout en travaillant sur un problème.

Maintenant, faite les exemples suivants.

Trouvez l'exemple 1 sur votre feuille de réponses.

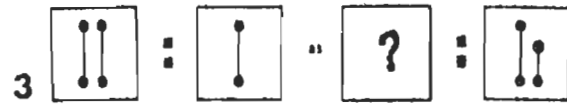


Trouvez l'exemple 2 sur votre feuille de réponses.



1 17,14,15,12,13,?

2 ZM : NY :: XO : ?



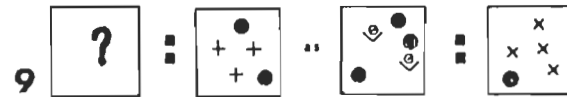
4

5 19,17,13,11,7,?

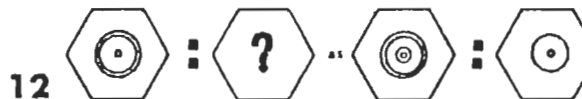
6 G,Q,H,P,I,?



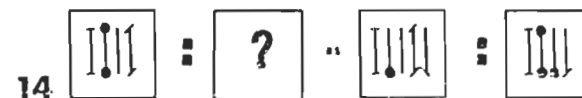
8 25,16,9,4,?



11 75% : 37.5% :: 50% : ?



13 DEARTH : EARTH :: DEAR : ?



1 a 10

b 11

c 9

d 7

2 a RZ

b OW

c WP

d PW



4

5 a 2

b 4

c 5

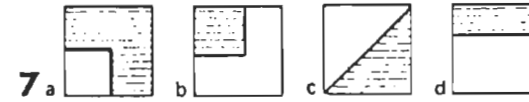
d 1

6 a N

b O

c J

d Q

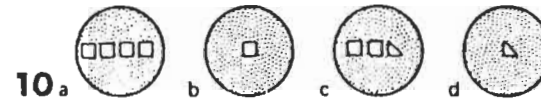


8 a 1

b 0

c 3

d 4



11 a $\frac{1}{2}\%$

b 25%

c 100%

d 15%



13 a GEAR

b FEAR

c SPEAR

d EAR



PROBLEMS

CHOICES

15 13, 1, 11, 2, 9, ?

16 M, K, I, G, E, ?

17 $\frac{3}{4} + \frac{1}{5} = \frac{3}{5} + \frac{1}{4} = ? = \frac{17}{20}$

18

19 19, 17, 13, 11, 7, ?



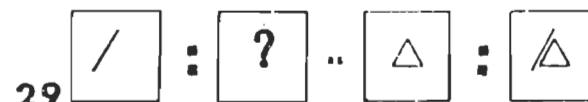
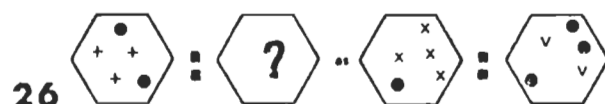
21 17, 14, 15, 12, 13, ?

22

23 25, 16, 9, 4, ?

24 ? : 25 :: 16 : 36

25 1, 3, 5, 7, 9, ?



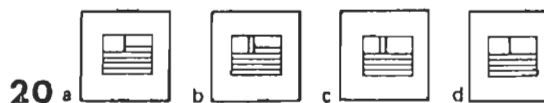
15 a 3 b 4 c 7 d 23

16 a B b D c C d F

17 a $\frac{17}{20}$ b $\frac{18}{20}$ c $\frac{13}{20}$ d $\frac{19}{20}$

18

19 a 1 b 3 c 5 d 2



21 a 10 b 11 c 9 d 7

22

23 a 5 b 4 c 0 d 1

24 a 5 b 4 c 9 d 15

25 a 0 b 13 c 11 d 12



A B C D E F G H I J K L M N O P Q R S T U V W X Y Z 0 1 2 3 4 5 6 7 8 9

PROBLEMS

CHOICES

30 $8 \div 27 \approx 64 \div ?$

31 IF $2 \times 2 = 4$ and $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ then $2\frac{1}{2} \times 2\frac{1}{2} = ?$

32 $2.20 \div 2.30 \approx ? \div 2.10$

33 $? \div \text{RITE} \approx \text{SIGHT} \div \text{SITE}$

34 $153 \div 264 \approx ? \div 486$

35  \div  \approx  \div 

36 $4.5 \div 9 \approx ? \div 9$





37 $K, N, J, O, I, ?$

38 $5.9 \div ? \approx 0.9 \div 1.0$

39 $AEC \div ERDA \approx OSS \div ?$

40 $2, 4 \div 3, 5 \approx ? \div 5, 7$

41  \div  \approx  \div 

42  \div  \approx  \div 

43  \div  \approx  \div 

30 a 54 b 81 c 83 d 629

31 a $6\frac{1}{4}$ b $4\frac{1}{4}$ c $\frac{25}{4}$ d 5

32 a 2.04 b 2.00 c 2.20 d 2.02

33 a RIGHT b WROTE c ROTE d WRITE

34 a 573 b 375 c 357 d 537

35 a  b  c  d 

36 a 2,6 b 2,5 c 3,6 d 6,3





37 a Q b P c W d H

38 a 5.0 b 5.8 c 6.0 d 6.9

39 a FBI b APO c HEW d CIA

40 a 2,6 b 6,8 c 4,6 d 4,8

41 a  b  c  d 

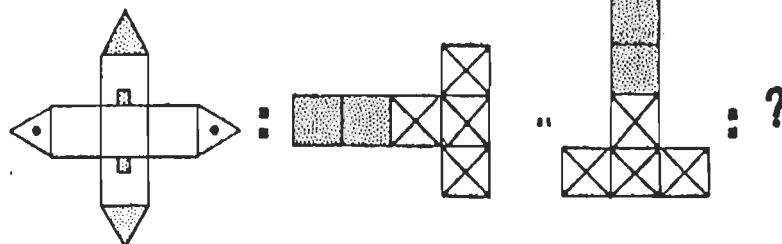
42 a  b  c  d 

43 a  b  c  d 

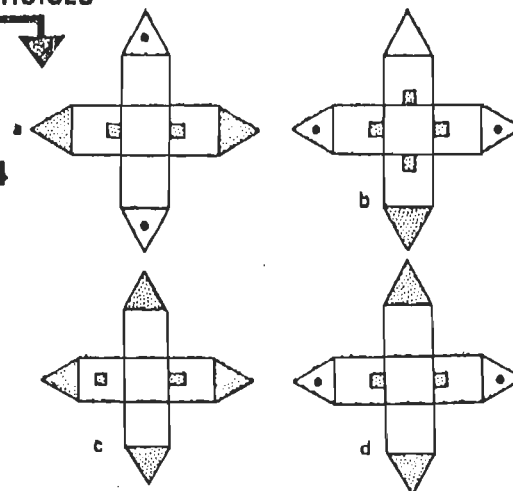
PROBLEMS

CHOICES

44



44



45 DEF, EFD, FDE, LMN, ?

45 a NLM b NML c MLN d MNL

46 GOAL : ? :: LIAR : LAIR

46 a LOGA b GOOL c GAOL d POST



47 a

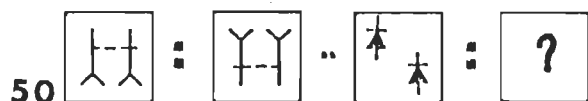


48 ABC : IHG :: ? : JKL

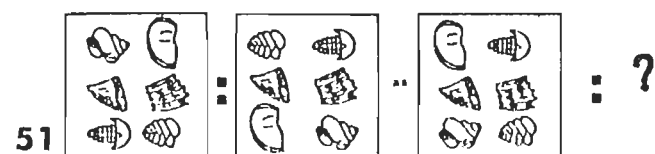
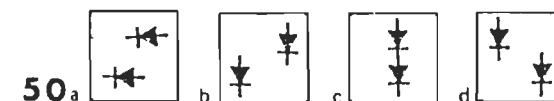
48 a FED b EFD c DEF d CDE



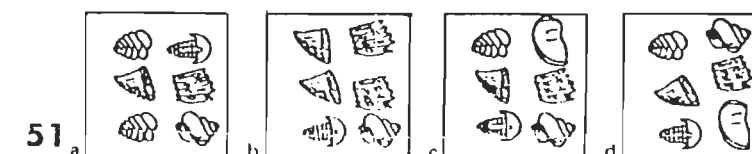
49 a



50 a



51 a

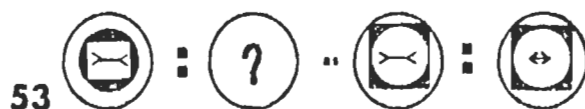


A B C D E F G H I J K L M N O P Q R S T U V W X Y Z 0 1 2 3 4 5 6 7 8 9

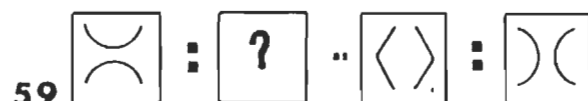
PROBLEMS

CHOICES

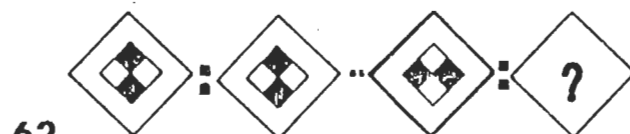
52

54 $XN : HL :: OA : ?$ 56 $\frac{1}{3} : 33.3\% :: ? : 25\%$ 

58 93, 39, 54, 45, 51, ?



60

61 FACT : ? :: $\sqrt{1}$: $\sqrt{-1}$ 

63

64 $\frac{1}{3} - \frac{1}{4} : \frac{1}{12} :: \frac{1}{4} - \frac{1}{5} : ?$

52



54 a JW b IV c RO d ZS

56 a 50 b $\frac{1}{4}$ c $\frac{1}{5}$ d $\frac{1}{2}$ 

58 a 150 b 9 c 15 d 153



60

61 a FABLE b FACTUAL c FICTITIOUS d FORMIDABLE



63

64 a $-\frac{1}{20}$ b $\frac{1}{20}$ c +1 d -1

PROBLEMS

CHOICES

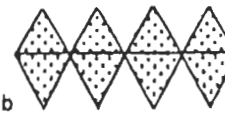
65



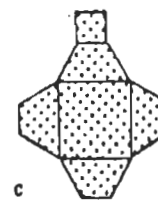
65.



b



c



d



66

IF $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{4}{8} + \frac{2}{8} + \frac{1}{8} = \frac{7}{8}$ then $\frac{1}{2} + \frac{1}{3} + \frac{1}{5} = ?$

66. a $1\frac{1}{30}$ b $3\frac{1}{30}$ c $\frac{16}{15}$ d $3\frac{1}{30}$

67

68

6% ■ 60% .. ? ■ 410

67

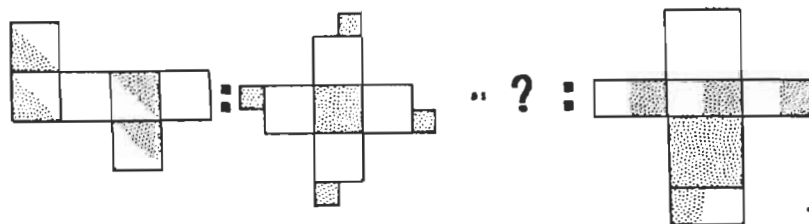
68. a 4.10

b 0.41

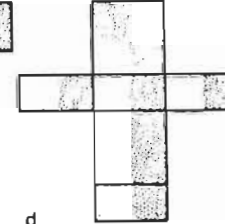
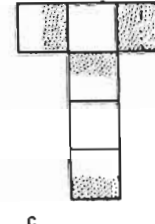
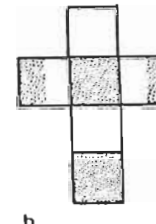
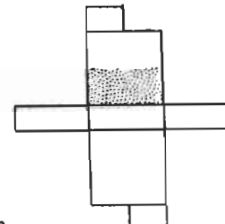
c 41.0

d 4100

69



69



70

71

20% ■ $\frac{1}{5}$.. 10% ■ ?

70. a

71. a $\frac{2}{10}$

b 90

c $\frac{10}{100}$ d $\frac{1}{10}$

72

73

IF $\frac{1}{2} - \frac{1}{4} - \frac{1}{8} = \frac{4}{8} - \frac{2}{8} - \frac{1}{8} = \frac{1}{8}$ then $\frac{1}{2} - \frac{1}{3} - \frac{1}{5} = ?$

72

73. a $\frac{1}{30}$

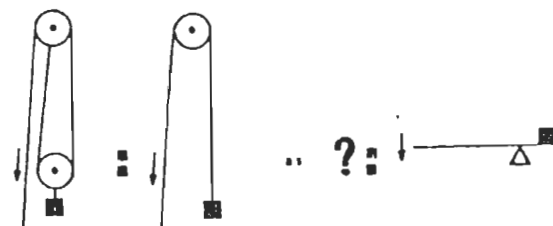
b 0

c $-\frac{1}{30}$

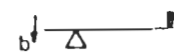
d -5

74

74



75. a



76

77

$\frac{1}{4} + \frac{1}{5} = \frac{5}{20} + \frac{4}{20} = \frac{9}{20}$.. ? ■ $\frac{7}{21} + \frac{3}{21}$

76

77. a $\frac{1}{7} + \frac{1}{3}$ b $\frac{3}{7} + \frac{3}{7}$ c $\frac{1}{3} + \frac{1}{7}$ d $\frac{3}{1} + \frac{7}{1}$

78

79

I, K, M, O, Q, ?

78

79. a T

b P

c R

d S

80

IF $3 \times 3 = 9$ and $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$ then $3\frac{1}{3} \times 3\frac{1}{3} = ?$

80. a $11\frac{1}{9}$ b $9\frac{1}{9}$ c $6\frac{2}{3}$ d $\frac{100}{9}$

Appendice B

Résultats complémentaires

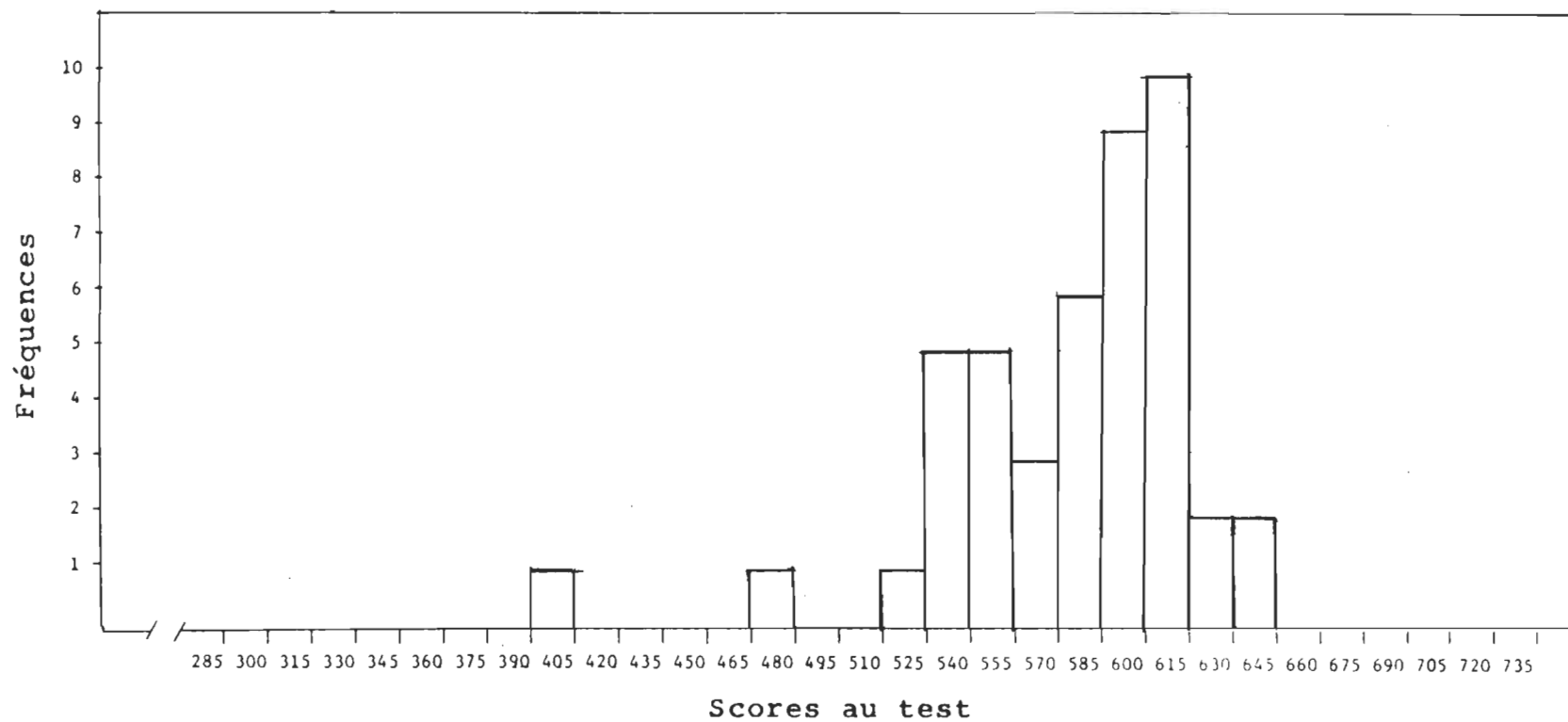


Fig. 4 - Histogramme de la distribution des scores au LAP pondéré sur 680 et corrigé selon la méthode AUC.

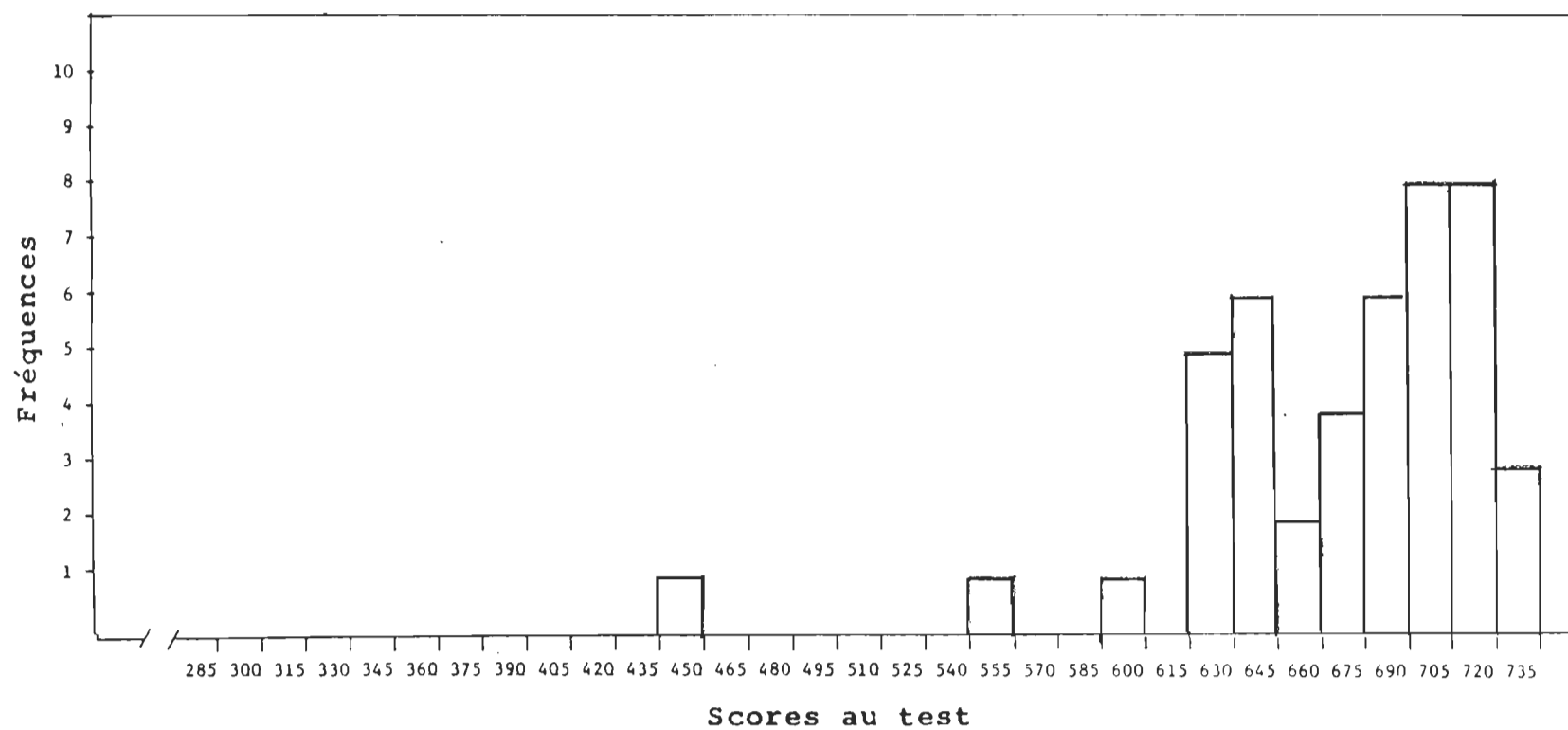


Fig. 5 - Histogramme de la distribution des scores au LAP pondéré sur 800 et corrigé selon la méthode AUC.

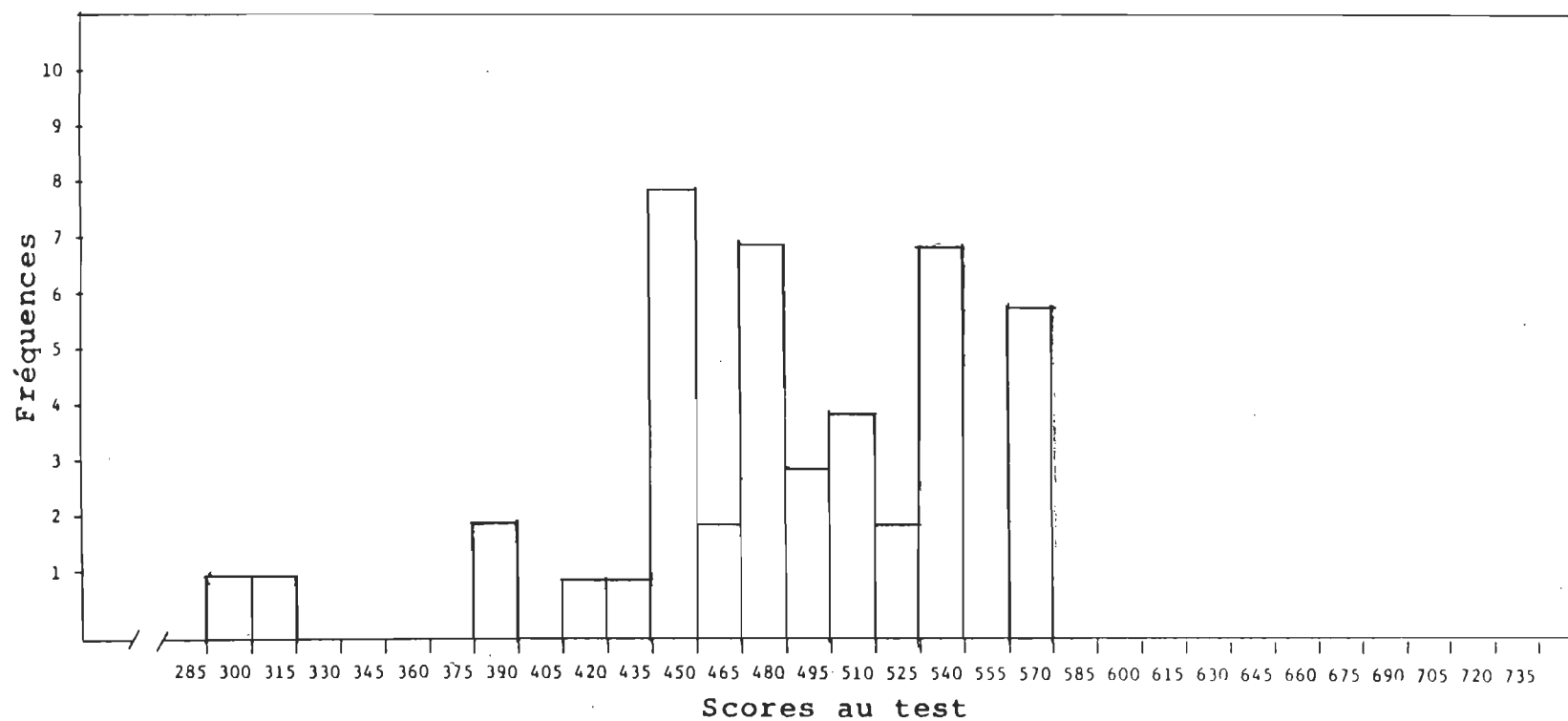


Fig. 6 - Histogramme de la distribution des scores au LAP pondéré sur 680 et corrigé selon la méthode INR.

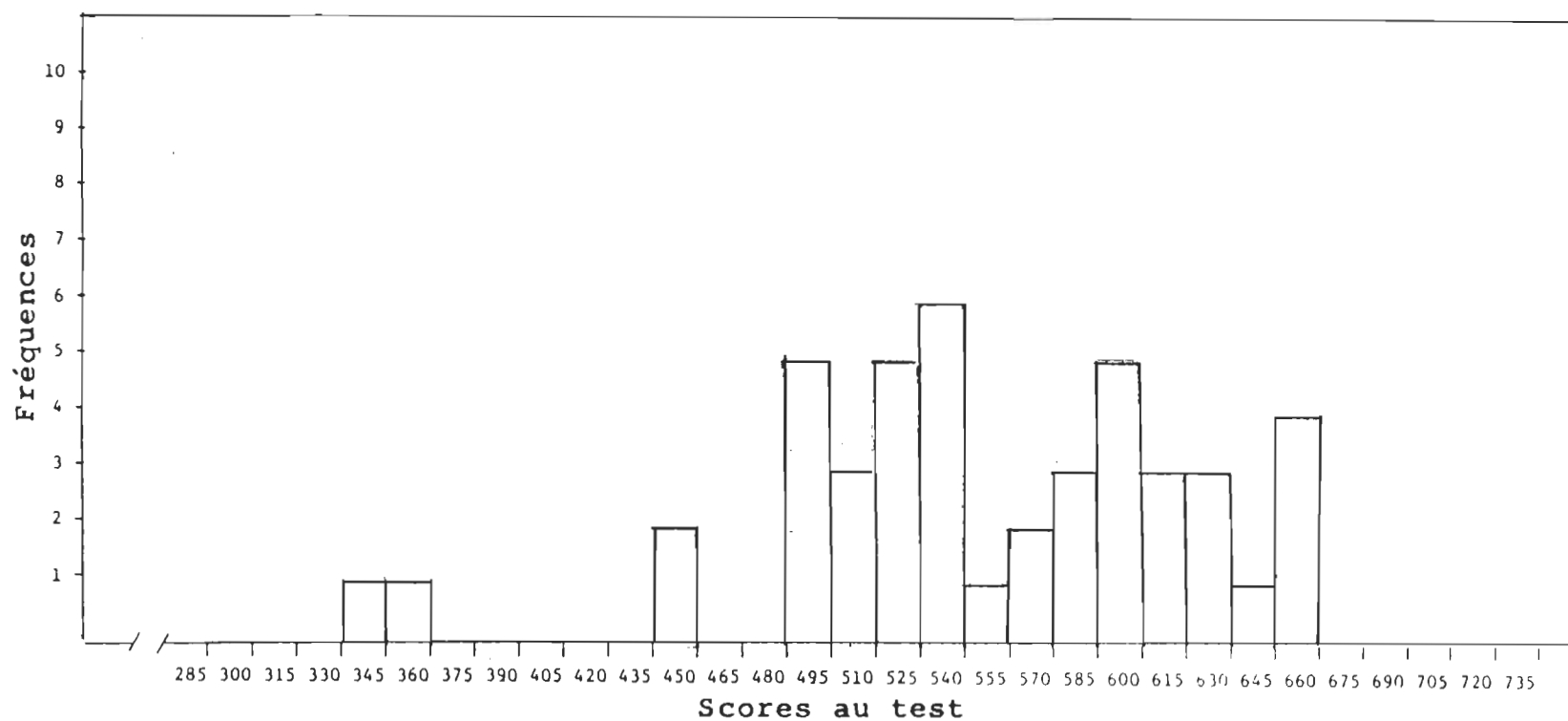


Fig. 7 - Histogramme de la distribution des scores au LAP pondéré sur 800 et corrigé selon la méthode INR.

Tableau 10

Distribution (%) des notes académiques et coefficients de validité prédictive (LAP-680, AUC) obtenus pour le groupe expérimental (N=45) en fonction des cours suivis

	%					
	A	B	C	D	E	r
Psy. de la perception I	12.5	4.5	27.5	7.5	7.5	0,391
Psy. de l'apprentissage	20.9	46.5	20.9	9.3	2.3	0,095
Histoire de la psy.	7.1	71.4	19.0	0.0	2.4	-0,057
Psychophysiologie I	19.0	45.2	21.4	11.9	2.4	0,380
Statistiques en psy.	29.4	44.1	8.8	8.8	8.8	0,314
Psy. dynamique I	13.2	76.3	10.5	0.0	0.0	0,080
Psy. des groupes I	30.0	47.5	22.5	0.0	0.0	-0,020
Psychophysiologie II	28.6	51.4	17.1	2.9	0.0	0,203
Psy. génétique I	26.5	47.0	26.5	0.0	0.0	0,189
Psy. sociale I	32.4	48.6	10.8	8.1	0.0	0,100

Tableau 11

Distribution (%) des notes académiques et coefficients de validité prédictive (LAP-680, AUC) obtenus pour les sujets féminins (N=24) en fonction des cours suivis

	%					
	A	B	C	D	E	r
Psy. de la perception I	19.0	38.1	28.6	4.8	9.5	0,636
Psy. de l'apprentissage	13.0	56.5	21.7	4.4	4.4	0,177
Histoire de la psy.	9.1	72.7	18.2	0.0	0.0	-0,081
Psychophysiologie I	13.0	56.6	13.0	13.0	4.3	0,466
Statistiques en psy.	16.7	50.0	5.5	16.7	11.1	0,456
Psy. dynamique I	22.2	77.8	0.0	0.0	0.0	0,037
Psy. des groupes I	23.8	47.6	28.6	0.0	0.0	-0,112
Psychophysiologie II	23.5	53.0	23.5	0.0	0.0	0,417
Psy. génétique I	26.3	52.6	21.1	0.0	0.0	0,324
Psy. sociale I	40.0	40.0	15.0	5.0	0.0	0,076

Tableau 12

Distribution (%) des notes académiques et coefficients de validité
prédictive (LAP-680, AUC) obtenus pour les sujets
masculins (N=21) en fonction des cours suivis

	%					
	A	B	C	D	E	r
Psy. de la perception I	5.2	52.6	26.3	10.5	5.2	-0,079
Psy. de l'apprentissage	30.0	35.0	20.0	15.0	0.0	-0,060
Histoire de la psy.	5.0	70.0	20.0	0.0	5.0	0,044
Psychophysiologie I	26.3	31.6	31.6	10.5	0.0	0,213
Statistiques en psy.	43.7	37.5	12.5	0.0	6.3	-0,128
Psy. dynamique I	5.0	75.0	20.0	0.0	0.0	0,488
Psychophysiologie II	33.3	50.0	11.1	5.6	0.0	-0,012
Psy. génétique I	26.6	40.0	33.3	0.0	0.0	0,056
Psy. sociale I	23.6	58.8	5.9	11.7	0.0	0,217

Remerciement

L'auteur désire exprimer sa reconnaissance à son directeur de thèse, monsieur Bruno Fabi, D. Ps., professeur au Département d'administration et d'économique à l'Université du Québec à Trois-Rivières, à qui il est redevable d'une assistance constante et éclairée.

Références

- ADAMS, G.S. (1964). Measurement and evaluation in education, psychology and guidance. Toronto: Holt, Rinehart and Winston.
- ALBRIGHT, L.E., GLENNON, J.R., SMITH, W.J. (1963). The use of psychological tests in industry. Cleveland, OH: Harvard Allen.
- ANASTASI, A. (1982). Psychological Testing, (5th ed.). New-York: Macmillan.
- ANGOFF, W.H. (1971). The College Board Admission Testing Program: A technical report on research and development activities relating to the Scholastic Aptitude Test and aptitude variables. Journal of educationnal measurement, 4, 143-149.
- ASTIN, A.W. (1971). Predicting academic performance in college: selectivity data for 2300 american colleges. New-York: Free Press.
- BABY, A. (1975). Quelques évidences oubliées à propos du testing en éducation. L'orientation professionnelle, 11, 101-113.
- BALDAUF, R.B. (1978). The validity of the Michigan Test of English Language Proficiency as a general measure of high school english achievement in American Samoa. Educational and psychological measurement, 38, 429-432.
- BEAN, A. (1975). The prediction of performance in an educational psychology master's degree program. Educational and psychological measurement, 35, 963-967.
- BEGIN, Y. (1980). Vers une réorientation de la psychométrie de l'aptitude: le point de vue de Bloom. Revue québécoise de psychologie, 1, 2-12.
- BELANGER, D. (1977). Les méthodes de recherche en sciences sociales. Editions HRW: Anjou.

- BELZIL, R. (1980). Etude de validité des tests d'aptitudes pour le programme du Certificat général en droit. Rapport inédit, Université de Montréal.
- BOUCHER, F. (1972). Rôle des tests d'Aptitudes aux études universitaires dans le processus d'admission à l'université. Délibérations et mémoires de la réunion annuelle du Service d'Admission au Collège et l'Université. Editions Marie Sivy.
- BOUCHER, F. (1975). Objectifs et fondements des Tests d'Aptitudes aux études universitaires. L'orientation professionnelle, 11, 25-34.
- BOUCHER, F. (1982). La prédiction du rendement scolaire à l'université par une mesure de la vitesse d'apprentissage. Revue québécoise de psychologie, 1982, 3, 18-26.
- BURKE, M.J. (1984). Validity generalization: a review and critique of the correlation model. Personnal psychology, 37, 93-116.
- CALLENDER, J.C., OSBURN, H.G. (1980). Development and test of a new model for validity generalization. Journal of applied psychology, 65, 543-558.
- CAMP, J., CLAWSON, T. (1979). The relationship between the Graduate Record Examinations Aptitude Test and graduate grade point average in a master of art in counseling program. Educational and psychological measurement, 39, 429-431.
- CAMPBELL, D.T. (1960). Recommendations for APA Test standards regarding construct, trait, or discriminant validity. American psychologist, 14, 546-553.
- CARDINAL, L. (1977). Etude sur la validité des Tests d'Aptitudes aux études universitaires selon une approche différentielle. Thèse de maîtrise inédite, Université de Montréal.
- CARTER, W.J. (1972). An analysis of empirically established hierarchies of academic fields of study bases on Graduate Record of Examination. Unpublished doctoral dissertation, University of Texas.

- CATTELL, R.B. (1967). La théorie de l'intelligence fluide et cristallisée, sa relation avec les tests "culture fair" et sa vérification chez les enfants de 9 à 12 ans. Revue de psychologie appliquée, 17, (3), 135-154.
- CATTELL, R.B., BARTON, K., DIELMAN, T.E. (1972). Prediction of school achievement for motivation, personality and ability measures. Psychological reports, 30, 35-43.
- CATTELL, R.B., BUTCHER, H.J. (1968). The prediction of achievement and creativity. New-York: Bobbs-Merrill.
- CHANSKY, N.M. (1964). A note on the grade point average in research. Educational and psychological measurement, 24, 96-99.
- CONRY, R., PLANT, W.T. (1965). WAIS and group test predictions of an academic success. Educational and psychological measurement, 25, 493-500.
- CORMIER, M. (1978). Le contingentement. L'orientation professionnelle, 14, 273-279.
- COSTIN, F. (1978). Do student ratings of college teachers predict student achievement? Teaching of psychology, 5, 86-88.
- CRONBACH, L.J. (1970). Essentials of psychological testing (3rd ed.). New-York, Harper.
- DAVIDSON, J.F. (1975). Academic interest rates and grade inflation. Educational record, 56, 122-125.
- DAYHAW, L.T. (1979). Manuel de statistique. Ottawa: Editions de l'Université d'Ottawa.
- DOLE, A.A., BAGGALEY, A.R. (1979). Prediction of performance in a doctoral education program by the Graduate Record Examinations and other measures. Educational and psychological measurement, 39, 421-427.
- DUBOIS, P.H. (1972). The College Board Scholastic Aptitude Test, in O.K. Buros (Ed.): The seventh mental measurements yearbook (pp. 646-648). Highland Park: Gryphon.

- ELLIOTT, H.A. (1972). Rapport du directeur associé. Délibérations et mémoires de la réunion annuelle du Service d'Admission au Collège et à l'Université. Editions Marie Sivyer.
- ENGELHART, M.D. (1965). The ACT assesment Program, in O.K. Buros (Ed.): The sixth mental measurements yearbook (pp. 2-7). Highland Park: Gryphon.
- ENO, L. (1978). Predicting achievement and the theory of fluid and crytallized intelligence. Psychological reports, 43, 847-852.
- ETAUGH, A.F., ETAUGH, C.F., HURD, D.E. (1972). Reliability of college grades and grade point averages: some implications for predictions of academic performance. Educational and psychological measurement, 32, 1045-1050.
- EVARTS, P. (1980). How to prepare for the American College Testing Program (A.C.T. Assesment). New-York: Mc Graw-Hill.
- EWEN, R.B. (1969). The GRE psychology test as an unobtrusive measure of motivation. Journal of applied psychology, 53, 383-387.
- FABI, B. (1983). A concurrent validity study of the Learning Ability Profile against college grade point average: some canadian data. Educational and psychological measurement, 43, 859-863.
- FERGUSON, G.A. (1976). Statistical analysis in psychology and education (4th ed.). New-York: Mc Graw-Hill.
- FINDLEY, W.G. (1965). The ACT Assesment Program, in O.K. Buros (Ed.): The sixth mental measurement yearbook (pp. 7-10). Highland Park: Gryphon.
- FRICKE, B.G. (1965). The College Entrance Examination Board Scholastic Aptitude Test in O.K. Buros (Ed.): The sixth mental measurements yearbook (pp. 975-988). Highland Park: Gryphon.
- GADBOIS, L., BRAULT, F. (1960). Rendement de deux tests en orientation scolaire. Montréal, Fédération des Collèges classiques.

- GARRETT, H.F. (1949). A review and interpretation of investigations of factors to scholastic success in colleges of arts and sciences and teachers colleges. Journal of experimental education, 18, 91-138.
- GAUDRY, R. (1982). Rationaliser l'enseignement supérieur. Affaires universitaires, 23, 2-3.
- GHISELLI, E.E. (1966). The validity of occupational aptitude tests. New-York: Wiley.
- GOLDMAN, R.D., SLAUGHTER, R.E. (1976). Why college grade point average is difficult to predict. Journal of education and psychology, 68, 9-14.
- GOULET, O. (1982). L'Université via ses différents programmes de formation. Service de Psychologie et d'orientation du Collège de Sherbrooke.
- GUEDON, M.C. (1979). Etude de la validité des critères de sélection utilisés à la faculté de Médecine dentaire en 1978-79. Rapport inédit, Université de Montréal.
- GUION, R.M. (1965). Personnal Testing. New-York: Mc Graw-Hill.
- HACKMAN, J.R., WIGGINS, N., BASS, A.R. (1970). Prediction of long term success in doctoral work in psychology. Educational and psychological measurement, 20, 365-374.
- HANNA, G.S. (1977). A study of reliability and validity effects of total and partial immediate feedback in multiple-choice testing. Journal of educational measurement, 14, 1-7.
- HARGADON, F. (1981). Tests and college admissions. American psychologist, 36, 1112-1119.
- HARTNETT, J.L., FELDMESSER, R.A. (1980). College admissions testing and the myth of selectivity: unresolved questions and needed research. A.A.H.E. Bulletin, 32, 3-6.
- HARTNETT, R.T., WILLINGHAM, W.W. (1980). The criterion problem: what measure of success in graduate education? Applied psychological measurement, 4, 281-291.

- HARVEY, R.J. (1972). A comparaison of other test results with departmental examinations in New-Brunswick. Délibérations et mémoire de la réunion annuelle du Service de l'Admission au Collège et à l'Université. Editions Marie Sivyver.
- HENKEL, R.E. (1976). Tests of signifiance. Beverly Hills: Sage University Paper.
- HENNING, M. (1976). Learning Ability Profile. Denver, Colorado: Falcon Research and Development Co..
- HILLS, J.R. (1978). The ACT Assesment Program, in O.K. Buros (Ed.): The eight mental measurements yearbook (pp. 622-626). Highland Park: Gryphon.
- HIRSCHBERG, N., ITKIN, S. (1978). Graduate student success in psychology. American psychologist, 83, 1083-1093.
- HOUSTON, L.W. (1983). The comparative predictive validities of high school rank, the Ammons Quick Test, and two Scholastic Aptitude Test measures for a rample of black female college students. Educational and psychological measurement, 43, 1123-1126.
- JONES, W.P. (1970). Sex differences in academic prediction. Measurement and evaluation in guidance, 3, 88-91.
- KERLINGER, F.N. (1973). Foundations of behavioral research. New-York: Holt, Rinehart and Winston.
- KHAN, S.B. (1973). Sex differences in predictability of academic achievement. Measurement and evaluation in guidance, 6, 88-92.
- LAMONTAGNE, L. (1972). Rapport du directeur général. Délibérations et mémoires de la réunion annuelle du Service de l'Admission au Collège et à l'Université. Editions Marie Sivyver.
- LAVIN, D.E. (1965). The prediction of academic performance. New-York: Russel Sage Foundation.
- LAVOIE, B. (1973). Tests SACU vs résultats académiques. L'orientation professionnelle, 9, 229-232.

- LINN, R.L. (1982). Admissions testing on trial. American psychologist, 37, 279-291.
- LINN, R.L., DUNBAR, S.B. (1982). Predictive validity of admissions measures: corrections for selection on several variables. Journal of college student personal, 23, 222-226.
- LINN, R.L., HARNISCH, D.L., DUNBAR, S.B. (1981). Validity generalization and situational specificity: an analysis of the prediction of first-year grades in law school. Applied psychological measurement, 5, 281-289.
- LITTLEPAGE, G.E., BRACC, D.M., RUST, J.P. (1978). Relations between admission criteria, academic performance and professional performance. Teaching of psychology, 5, 16-20.
- MERENDA, P.F., REILLY, R. (1971). Validity of selection criteria in determining success of graduate students in psychology. Psychological reports, 28, 259-266.
- MICHAEL, J.J., NADSON, J.S., MICHAEL, W.B. (1983). The prediction of academic achievement in graduate study in education. Educational and psychological measurement, 43, 1133-1139.
- MOTOYAMA, T., WOLINS, L. (1980). On an indicator of good prediction. Educational and psychological measurement, 40, 939-942.
- MUNDAY, L.A. (1970). Factors influencing the predictability of college grades. American educational research journal, 7, 99-107.
- NAIRN, A. (1980). The reign of ETS: the corporation that makes up minds. Washington, D.C.: Author.
- NIE, N.H., HADLAI HULL, C., JENKINS, J.C., STEINGRENNER, K., BENT, D.H. (1975). Statistical package for the social sciences. New-York: Mc Graw-Hill.
- OMIZO, M.M., RIVERA, E. (1980). Self-actualization measures as predictors of ability in facilitative communication among counselor trainees. Educational and psychological measurement, 40, 451-456.

- PASSONS, W.R. (1967). Predictive validities of the ACT, SAT and high school grades for first semester GPA and freshman courses. Educational and psychological measurement, 15, 487-490.
- PEARLMAN, K., SCHMIDT, F.L., HUNTER, J.E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of applied psychology, 65, 373-406.
- PORTER, A.L., WOLFE, D. (1975). Utility of the doctoral dissertation. American psychologist, 30, 1054-1061.
- QUERESHI, M.Y., MILLER, J.M. (1970). The comparability of the WAIS, WISC and WB II. Journal of educational measurement, 7, 105-111.
- ROY, R. (1972). Recherche SACU vs Résultats scolaires. L'orientation professionnelle, 8, 229-232.
- SELLTIZ, C., WRIGHTSMAN, L.S., COOK, S.W. (1977). Research methods in social relations. (third edition). Holt, Rinehart and Winston.
- SCHMIDT, F.L., HUNTER, J.E. (1977). Development of a general solution to the problem of validity generalization. Journal of applied psychology, 66, 261-273.
- SCHMIDT, F.L., HUNTER, J.E., PEARLMAN, K. (1982). Progress in validity generalization: comments on Calender and Osburn and further developments. Journal of applied psychology, 67, 835-845.
- SCHMIDT, F.L., KAPLAN, L.B. (1971). Composite vs multiple criteria: a review and resolution of the controversy. Personnel psychology, 24, 419-434.
- SCHRADER, W.B. (1971). The predictive validity of College Board admissions tests. New-York: college Entrance Examination Board.
- SEASHORE, H.G. (1962). Women are more predictable than men. Journal of counseling psychology, 9, 261-270.
- SINGLETON, R., SMITH, E.R. (1978). Does grade inflation decrease the reliability of grades? Journal of educational measurement, 15, 37-41.

- SLACK, W.V., PORTER, D. (1980). The Scholastic Aptitude Test: a critical appraisal. Harvard educational review, 50, 154-175.
- STANLEY, J.C. (1967). Further evidence via the analysis of variance that women are more predictable academically than men. Ontario journal of educational research, 10, 49-56.
- STATISTIQUE CANADA (1982). Statistique de l'enseignement - Estimations, 1982-1983. Ottawa: Ministère des Approvisionnements et Services.
- THACKER, J.A., WILLIAMS, R.F. (1974). The relationship of the Graduate Record Examination to grade point average and success in graduate school. Educational and psychological measurement, 34, 939-944.
- TRYON, W.W. (1979). The test-trait fallacy. American psychologist, 34, 402-406.
- VOYER, J.P. (1972). Au sujet de la validation d'un examen d'admission. Délibérations et mémoires de la réunion annuelle du Service d'Admission au Collège et à l'Université. Editions Marie Sivyer.
- WALLACE, W.L. (1972). The college Board Scholastic Aptitude Test, in O.K. Buros (Ed.): The seventh mental measurements yearbook (pp. 648-650). Highland Park: Gryphon.
- WHITE, E.M. (1975). Sometimes an A is really an F. Chronicle of higher education, 9, 24.
- WILLINGHAM, W.W. (1974). Predicting success in graduate education. Science, 183, 273-278.
- ZELEZNIK, C., HOJAT, M., VELOSKI, J. (1983). Long-range predictive and differential validities of the Scholastic Aptitude Test in medical school. Educational and psychological measurement, 43, 223-230.
- ZUNG, W.W., GIANTURCO, J. (1968). Further validation of the Ohio literacy Test: correlation with the Wechsler Adult Intelligence Scale and grade achieved in school. Journal of clinical psychology, 24, 197-198.